



HAL
open science

Measuring member contribution impact in an online community

Hirotoishi Takeda

► **To cite this version:**

Hirotoishi Takeda. Measuring member contribution impact in an online community. Business administration. Université Paris Dauphine - Paris IX, 2015. English. NNT : 2015PA090034 . tel-01250810

HAL Id: tel-01250810

<https://theses.hal.science/tel-01250810>

Submitted on 5 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE PARIS-DAUPHINE

ECOLE DOCTORALE

Dauphine Recherche en Management

THESE

pour l'obtention du titre de

DOCTEUR EN SCIENCES DE GESTION
(Arrêté du 7 août 2006)

Présentée et soutenue publiquement par

Hirotooshi Takeda

Le 29 septembre 2015

**MEASURING MEMBER CONTRIBUTION IMPACT IN AN ONLINE
COMMUNITY**

JURY

- Directeur de thèse :** **Michel KALIKA**
Professeur, IAE Lyon, Université Jean Moulin
- Rapporteurs :** **Alain CUCCHI**
Professeur, IAE, Université de la Réunion
Jessie PALLUD
Professeur, EM Strasbourg Business School,
Université de Strasbourg
- Suffragants :** **Christophe ELIE-DIT-COSAQUE**
Professeur, IAE, Université de Lorraine
Frantz Rowe
Professeur à Université de Nantes
Duane Truex III
Professeur à Georgia State University

Remerciements

Pursuing a doctoral degree at Dauphine was something that was on the back of my mind since completing my degree at GSU. But deciding to pursue a doctoral degree in general was something that I was not able to pursue until I knew I had the support that I needed. While achieving any degree requires support from colleagues, friends, and family, this was truly the case with my journey to pursue a doctoral degree.

Within academia I would like to first thank Ms. Florence Parent for her help every time there were issues to register or to put in applications at Dauphine. I have learned the hard way how difficult and important it is to get the right person to help with the proper documents. I would also like to thank my committee members Dr. Michel Kalika, Dr. Frantz Rowe, Dr. Christophe Elie-Dit-Cosaque, Dr. Jessie Pallud, and Dr. Alain Cucchi as well as advisors Dr. François-Xavier de Vaujany and Dr. Isabelle Walsh. I would also like to thank my advisors at GSU, Dr. Duane Truex, Dr. Michael Gallivan, Dr. Bala Ramesh and Dr. Richard Baskerville. I would also like to thank Dr. Roy Johnson for allowing me the opportunity to start my Ph.D. career at GSU and the staff including Libby and Adenike. I would also like to thank various professors that advised and taught courses that I took during my studies. I would also like to thank my colleagues and co-authors of various papers that make up my dissertation. These include Dr. Richard Vidgen and Dr. Brett Young, Dr. Robert Sainsbury.

I would like to thank my Ph.D. student colleagues at GSU including Dr. Yi Ding, Dr. Alina Dulipovici, Dr. Jong Woo Kim, Dr. Chongwoo Park, Dr. Yide Shen, Dr. Sweta Sneha, Dr. Anthony Vance, Dr. Jijie Wang, and Dr. Jack Zheng. I would also like to thank my Ph.D. student colleagues at the University of Paris Dauphine, Dr. Anouck Adrot, Dr. Mathieu Chauvet, Dr. Antoine Harfouche, Dr. Hamed Motaghi, and Dr. Gaetan Mourmant.

I would also like to thank my friends in France especially those at Nantes University, Dr. Frantz Rowe, Dr. François de Corbière, Dr. Johanna Habib, Dr. Nicolas Antiaume, and Dr. Daniel Thiel. I also want to thank those on the Nantes Women's hockey team, especially Leslie, Julia, Sara, Cloé, Audrey, Sapphire, Aurore, Stefani, Béatrice, Nathalie, and Dave Lapierre. Also those in Paris, Mr. and Mrs Urata, Mr. Yamano Norihiko, and including Quicu who was an unlikely friend during my time in Paris.

I would also like to extend my thanks to my GAP program colleagues Dr. Stephen Du and Dr. Nannette Napier, whom made my transition to GSU much more tolerable. I would also like to extend a special thanks to Dr. Michael Cuellar, whom has been my colleague as a student, when we first started the GSU GAP program together, as a co-author, which we have worked on various papers for over 5 years, and as colleagues at North Carolina Central University.

I would like to thank my advisor Dr. Duane Truex, whom has been my guide in my travels through my dissertation process. I would also like to thank his family Dr. Cherie Long, Sean-Ryan, and Colin for putting up with my busy schedule.

The studies of a doctoral student and the writing of a dissertation is an extremely frustrating and somewhat lonely endeavor. This frustration and loneliness of the dissertation process has caused me to rely on my personal life to keep me sane most of the time. I would like to thank Coach Lou Solo and the late Coach Greg Stathis for their support over the years. I would also like to thank the many

referee friends that have showed concern through my academic process. I also want to thank the many Georgia Tech Hockey Team players that I have been privileged to have coached and acquainted with. They were the ones that kept me young at heart. I would also like to thank my friends, including those that I have met through refereeing hockey. Thank you Rob, Nathan, Barry, Christopher, Jeremy, Jason, Chuck, Roger, Bob, Billy, Scott, Alan, Suzanne, Keith, Brian, Eric, Mark, John, Eric, Mitch, Gary, Leo, Tommy, Vicki, Kim, Stephanie, Erik, Bobby, Andrew, Rhett, Ken, Brett, David, Andy, Brad, Joe, Kelly, Sammi, Sean, Jeff, Shawn, Tim, Kevin, Don, and those that passed Edd Baker and Scott Olson.

I would like to especially thank my family, my siblings, Eiji and Miyuki, as well as my grandparents, uncle Shinichi, and especially my parents Kazuko and Toshihide. My parents have been patient and understanding of the stresses of the Ph.D. process even when they did not quite fully understand the academic life. They have been supportive through and beyond the whole process.

I would also like to thank Kitty, Currie, and Nala for keeping me grounded. Finally I would like to thank my wife Sara Crabtree who has been my confidant throughout this process. You were the one that encouraged me to come to GSU and pursue a Ph.D. degree in the first place. You have been my partner in life, my co-author, and my editor throughout the process. Thank you cannot say enough of what you have meant to me.

Thank you y'all.

L'Université n'entend donner aucune approbation, ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.

Sommaire

Remerciements	3
Sommaire	6
Chapitre 1 Introduction générale (FR).....	7
Chapitre 2 General Introduction (EN).....	28
Chapitre 3 Investigating the Impact of Offline Events on Group Development in an Online Sports Community	47
Chapitre 4 Online Community User Influence: A Study Using User Status	57
Chapitre 5 What Drives User Contribution in an Online Community ? A Study in Contributor Influence and User Status	66
Chapitre 6 Online Community Influence: A Study Using the Hirsch Metric and Social Network Analysis	79
Chapitre 7 Measuring New Member Contribution in an Online Community: A Preferential Attachment Approach	91
Chapitre 8 Measuring Member Contribution Impact in an Online Community: A New Hirsch Index Inspired Approach.....	103
Chapitre 9 Conclusion générale (FR)	122
Chapitre 10 General Conclusions (EN).....	132
References.....	141
Table des matières	148
Table des figures.....	155
Table des tableaux.....	156

Chapitre 1

Introduction générale (FR)

La communauté en ligne (CEL) est une forme très répandue de transfert de connaissances spécialisées, où des usagers géographiquement dispersés peuvent constituer une communauté en partageant des idées, envoyant et affichant des messages, débattant de sujets et nouant des amitiés en ligne. Un des problèmes avec ces CEL est leur durabilité, car leur apparition et leur croissance initiale sont suivies d'une phase de stagnation où les usagers cessent d'afficher des commentaires, ce qui mène la communauté à mourir par manque d'activité. Tenter de prolonger la phase dynamique de croissance d'une CEL est un sujet pertinent pour tout administrateur de CEL. Une façon de maintenir le dynamisme d'une CEL est d'encourager les contributions.

Ce courant de recherche se penche sur la façon dont les CEL peuvent prolonger leur phase dynamique, en considérant différents aspects, en particulier les mesures des contributions des usagers et la manière dont les nouveaux usagers d'une CEL se comportent. Je propose d'utiliser différentes mesures pour évaluer les contributions des usagers. Une des mesures pour identifier les contributeurs très actifs est une mesure bibliométrique non-invasive basée sur l'indice de Hirsch. Un autre aspect de cette recherche concerne la façon dont les nouveaux usagers se comportent et comment cela peut être expliqué par l'attachement préférentiel.

1.1. Introduction

Les communautés en ligne continuent à croître en popularité et sophistication, permettant à des membres géographiquement dispersés qui partagent des intérêts similaires de communiquer et d'échanger des idées, des opinions et des contenus multimédia. Ces CEL offrent à des personnes partageant des intérêts des moyens de participer activement au partage d'information, à des débats sur des sujets pertinents (ou non), à la création de liens sociaux virtuels, et à la construction d'un sentiment d'appartenance à une CEL dynamique. Les CEL sont souvent utilisées comme sources de savoir et dépôt d'informations que les membres créent à partir de leurs expériences dans la vie réelle. Mais pour que ces communautés existent, leurs membres doivent y participer activement. Leurs contributions doivent être également pertinentes, correctes et mises à jour.

Une communauté en ligne est définie comme un « agrégat d'individus ou de partenaires commerciaux qui sont en relation autour d'un intérêt commun, où la technologie, au moins partiellement, permet une interaction guidée par un certain nombre de protocoles ou de normes » (Porter 2004). Parfois, on appelle aussi une CEL une communauté virtuelle ou une communauté Internet. Ces CEL peuvent être des lieux de rassemblement pour les gens ayant des intérêts similaires où ils peuvent partager des informations, parler de leurs expériences, ou tout simplement se retrouver entre amis pour bavarder.

Bien que des CEL nouvelles apparaissent tout le temps, j'en ai vu un grand nombre disparaître. Les CEL qui ont perdu leur importance se trouvent souvent être des groupes d'intérêt particulier de Yahoo ou Google. Ces groupes peuvent être connus, car ils ont tendance à être des forums où des questions posées ont déclenché de vifs débats et discussions, mais ont été fermés ou sont devenus inactifs en raison d'un manque de participation ou de pertinence. D'autres CEL, telles que Second Life et Myspace, maintenant moins répandues, peuvent être considérées comme des sites de réseaux sociaux par leurs caractéristiques et certains éléments typiques des CEL. Ces sites de réseaux sociaux donnent la possibilité d'afficher et partager des informations entre usagers, ce qui est une caractéristique propre à toute CEL. Des CEL dynamiques et florissantes telles que Tripadvisor, Flyertalk et Rotten Tomatoes existent également. Chacun de ces sites propose des moyens de signaler des critiques de produits—des hôtels pour Tripadvisor ou des films pour Rotten Tomatoes ; les usagers peuvent commenter sur les commentaires, pour en indiquer le niveau d'utilité aux autres usagers.

Parfois, une CEL est liée à une entité qui en elle-même peut gagner en popularité, mais la perdre au fil du temps. Il y a beaucoup de CEL dédiées au jeu ou au sport où un produit ou une franchise peut réussir pendant un certain temps, mais échouer si le produit n'est pas mis à jour ou si l'équipe déménage (par exemple, Red Dead Redemption ou les Thrashers d'Atlanta). Il y a également de nombreux produits qui ont été tenus à jour et les CEL qui leur sont dédiées ont tendance à survivre (par exemple, Halo, World of Warcraft).

L'objectif du propriétaire d'une CEL est de maintenir le dynamisme de la communauté et d'engager ses membres. Pour cela, la communauté doit avoir des sujets et des discussions qui sont pertinentes et opportunes. Beaucoup de CEL tirent leurs revenus de la publicité et les recettes publicitaires sont liées à un mélange de consultations, de taux de clics sur les annonces, et d'achats initiés par les annonces. Un tel site est BigU, une communauté en ligne pour les supporters d'une équipe sportive universitaire qui sera présentée plus en détail ultérieurement.

Pour cette recherche, mon objectif général est de trouver des moyens de garder une CEL en place et en bon état de fonctionnement. Pour ce faire, la CEL doit être rentable afin de couvrir les frais d'organisation, de création de contenu et d'entretien. Pour que les revenus restent en hausse, la CEL doit maintenir une bonne base publicitaire, ce qui signifie, à tout le moins, que de nombreuses personnes doivent voir les annonces. Cependant, pour attirer les gens vers une CEL, son contenu doit être stimulant et pertinent. Par conséquent, la question de recherche spécifique pour ce projet est de savoir comment maintenir le dynamisme d'une communauté en ligne.

Dans le cadre de cette recherche, le terme « dynamisme » se réfère à l'interactivité dans la CEL. Le dynamisme est une mesure du nombre de messages affichés et de réactions et réponses qui suivent dans la CEL. Une CEL dynamique a des soumissions fréquentes et beaucoup d'interactions liées aux messages affichés. Quand une CEL perd son dynamisme, l'interaction devient stagnante, les messages affichés se font rares, et l'activité d'affichage et de réponse cesse. Pour garder son dynamisme, une CEL doit attirer beaucoup d'individus, et cela est directement lié au niveau d'intérêt du contenu créé par les usagers.

Un problème que j'ai rencontré est un manque de moyens pour identifier les usagers qui contribuent un contenu pertinent à une CEL. Le simple « nombre de messages affichés » est une mesure que beaucoup de CEL utilisent de nos jours pour montrer le niveau de contribution d'un usager. Malheureusement, ce chiffre mesure seulement la productivité, et en aucun cas l'impact, l'influence, et le nombre d'usagers qui lisent ces contributions. Ma recherche actuelle tente donc de trouver de meilleurs moyens de mesurer l'influence des contributeurs au sein d'une CEL.

Afin de mesurer l'influence et la productivité d'un usager et d'analyser les données d'une CEL, je propose une mesure globale utilisant l'analyse de réseau social et la bibliométrie (Takeda 2010). Cette mesure globale me permet d'identifier différents types d'usagers, ce qui était impossible en mesurant seulement « le nombre de messages affichés ».

1.2. Contexte

Je considérerai d'abord des CEL, en me concentrant en particulier sur leur durabilité et sur la manière de les dynamiser. Puis j'examinerai deux domaines principaux dont je tirerai les mesures, la bibliométrie et j'analyserai les réseaux sociaux.

1.2.1. Communautés en ligne (CEL)

Cette recherche porte sur la durabilité d'une CEL. Je m'intéresse aussi à la façon de maintenir la vitalité d'une CEL. Les CEL exigent des interactions, des contributions et des visites des membres de la communauté. Contrairement à un site de e-commerce vendant un produit tangible, une CEL nécessite une interaction importante des clients (qui sont les membres de la communauté) pour sa source de revenus, mais, dans la mesure où son contenu change rapidement, elle peut également avoir des clients qui reviennent régulièrement. Par exemple, dans la CEL que nous considérons plus loin, de nombreux usagers se connectent plusieurs fois par jour à cause du nouveau contenu et des nouvelles informations affichées tout au long de la journée. Tant que le contenu de la CEL continue à ajouter de la valeur pour le client, les propriétaires d'une CEL peuvent tirer un profit du service qu'elle offre.

Il existe de nombreuses études menées sur des CEL (Harper et al. 2007; Iriberry et Leroy 2009; Ren et Kraut 2007; Sassenberg et Postmes 2002), qui sont généralement constituées autour d'un sujet ou d'un intérêt. Les usagers deviennent membres d'une CEL afin d'obtenir plus d'informations ou partager des informations sur leur intérêt commun. La recherche a repéré différents types de CEL, par exemple des sujets tels que les films / le divertissement (Keen 1980, Wang et Fesenmaier 2003), les voyages (Wang et Fesenmaier 2004), les affaires (Wang et Fesenmaier 2003; Wang et al. 2008), la santé (Wang et Fesenmaier 2003), l'informatique et la science (Wang et Fesenmaier 2003; Zhang et al. 2007). Dans le domaine de la Science de l'Information, certaines études ont porté sur les systèmes de gestion des connaissances (Kankanhalli et al. 2005), d'autres sur le développement de logiciels, ciblant des CEL spécialisées dans le développement de logiciels open source (von Krogh et al. 2003; von Krogh et al. 2012), ainsi que dans le développement de logiciels de série (Holström et Henfridsson 2006).

La recherche actuelle sur la participation aux CEL est venue du domaine de recherche sur le partage de fichiers en pair-à-pair (P2P). Cette recherche avait tendance à considérer l'interaction contributeur et téléchargeur et comment la communauté fonctionnait à partir du travail des contributeurs (Golle et al. 2001). Plus récemment, la recherche cherche à définir une forme de système d'évaluation susceptible de mesurer l'impact des contributeurs et de leurs contributions. Par exemple, les usagers qui lisent les contributions pourraient être sollicités pour fournir du feedback. Les systèmes sont généralement mis en place pour reconnaître le contributeur ainsi que l'évaluateur sous forme d'une sorte de compensation (points pour contribution, monnaie virtuelle, etc.).

La réussite et la durabilité d'une CEL dépendent grandement de la participation des membres de la communauté. La qualité d'une CEL vient seulement du contenu créé par ses membres (Assmann et al. 2009; Cothrel et al. 1999). Les CEL ont tendance à avoir un cycle de vie limité où le dynamisme et la pertinence d'origine se maintiennent un moment avant de sombrer quand la communauté fournit un minimum de messages et une faible interaction sur une longue période (Iriberry et al. 2009). Les CEL qui s'éteignent ont des contributions sans pertinence dont l'information n'a plus cours, ce qui pousse les membres à cesser de visiter la CEL. La stagnation diminue le nombre de nouvelles contributions qui crée à son tour un cercle vicieux de comportement stagnant par les membres de la communauté. Même si des visiteurs viennent en nombre sur une CEL, les visites seules ne garantissent pas son dynamisme puisque l'interaction de la communauté n'est pas statique (Mousavidin et al. 2009). Certaines études de CEL traitent de leur décadence (Gainer et al. 2003; Keen 1980; Koh et al. 2007; Rothaermel et Sugiyama 2001; Trevino 1999). Une CEL peut attirer des visiteurs tant qu'un contenu nouveau et intéressant est disponible. Cependant, au fur et à mesure que la fréquence des nouvelles contributions diminue, les visites des membres de la communauté s'espacent. Une fois qu'un CEL perd son élan, essayer de faire revenir les anciens visiteurs devient beaucoup plus difficile. De nombreuses études ont tenté de résoudre ce problème en essayant de déterminer les raisons de la stagnation et d'identifier les moyens d'améliorer les contributions à une CEL (Gainer et al. 2003; Preece 2001; Ren et al. 2007; Rothaermel et Sugiyama 2001; Trevino 1999). Il existe donc un besoin pratique d'assurer que le dynamisme d'une CEL ne soit pas perdu. Le défi pour les propriétaires des CEL est d'essayer d'étendre la croissance initiale ou la période dynamique intermédiaire aussi longtemps que possible pour maintenir leur succès. Alors que de nombreux articles ont été écrits sur de grandes CEL telles que Myspace, Facebook, ebay, j'ai constaté un manque de recherche sur la durabilité de CEL plus petites (Clavio 2008; Shen et al. 2006).

Il existe aussi des recherches sur l'identification des usagers qui contribuent aux CEL. De nombreuses CEL utilisent le « statut de l'utilisateur », dérivé du nombre de messages affichés par un usager, pour mesurer l'activité des usagers. Typiquement, le « statut de l'utilisateur » est un critère utilisé pour désigner un usager qui a contribué une certaine quantité de messages ou a été actif pendant un certain temps. Par exemple, dans une CEL de sport, un usager peut devenir « remplaçant », « novice », « habitué », « vedette » en intervenant, respectivement, 5, 10, 100, et 1000 fois. Le statut de l'utilisateur dans les CEL était un domaine de recherche dans le passé. Ce domaine est basé sur le capital social (Ganley et al. 2006; Law et al. 2008), la présence sociale (Shen et al. 2009; Shen et al. 2006), l'identification sociale (Shen et al. 2008; Zheng et al. 2007), et sur la théorie de l'influence sociale (Assmann et al. 2009; Cheng et al. 2006). Certaines de ces études tentent de faire le lien entre les

usagers performants et le dynamisme et, par conséquent, le succès des CEL (Assmann et al. 2009; Cheng et al. 2006).

Ma thèse est que le critère de « statut de l'utilisateur » est seulement un outil de calcul et en aucun cas une mesure de la façon dont les membres de la communauté reçoivent les messages. On peut afficher des milliers de messages, mais si ces messages ne sont lus par personne dans la CEL, ils ne contribuent en rien à son succès et à sa vitalité.

Certaines études ont évalué les raisons d'une baisse d'activité dans une communauté et la façon d'aider les communautés stagnantes à devenir plus actives (Harper et al. 2007). Plusieurs études ont montré que les « observateurs passifs » existent dans toutes les CEL et que cette forme de « paresse sociale » leur est préjudiciable (Shiue et al. 2010). Certaines études concluent que la « paresse sociale » ou « l'observation passive » contribuent à leur disparition (Dantzig 2002; Gainer et al. 2003; Keen 1980; Rothaermel et Sugiyama 2001). La « paresse sociale » est « la tendance des individus à faire moins d'effort dans un projet collectif que dans un projet individuel » (Shiue et al. 2010 p. 769). « L'observation passive » consiste à lire « des messages dans une discussion sur Internet ... sans y contribuer » (Merriam-Webster Online Dictionary 2010). Le manque de participation peut indiquer de l'insatisfaction et une absence de raison impérieuse de devenir actif. Les « observateurs passifs » peuvent être moins satisfaits que les contributeurs actifs ou ne veulent pas ajouter à la confusion de forums de discussion déjà très chargés (Preece 2004).

Un domaine de recherche a essayé de changer les « observateurs passifs » en « contributeurs » et tenté d'identifier la motivation des contributeurs à participer. Leurs raisons ont été identifiées comme un désir d'améliorer leur réputation en ligne (Kankanhalli et al. 2005; Ma et Agarwal 2014; Ren et al. 2012; Wasko et Faraj 2005) et / ou simplement de contribuer pour la satisfaction d'aider les autres (Kankanhalli et al. 2005; Wasko et Faraj 2005). D'autres études ont examiné spécifiquement le mécanisme de feedback qui informe les contributeurs de leur gain en réputation dans une CEL (Burke et al. 2009; Cheshire et Antin 2008; Jung et al. 2010).

1.2.2 Méthodes invasives

Deux types de méthodes sont fréquemment présentes dans les publications: sondage et évaluation. Les deux méthodes nécessitent une certaine forme d'intervention de la part de l'utilisateur. Les études de feedback ont un mécanisme directement intégré dans la CEL. Généralement, ces systèmes comprennent une fonction permettant aux usagers d'« évaluer » le travail d'un autre usager ou de

l'unité de mesure cible du site. Par exemple, certaines CEL permettent à un usager d'évaluer les contributions des autres. Si une question est posée par un usager et que plusieurs personnes y répondent avec plus ou moins de détail, un usager est en mesure d'évaluer chacune des réponses données. Alors qu'il est possible d'avoir un système de notation multi-niveau, en général, ces types de sites permettent seulement le système Booléen de notation « Oui / Non » (Gainer et al. 2003; Keen 1980; Trevino 1999). Un exemple d'étude menée sur une unité de mesure visait un site de critique de film (Keen 1980). D'autres visiteurs du site donnent leur feedback par une note globale et des commentaires. Ces systèmes de feedback permettent généralement aux usagers de voir d'autres évaluations qui leur fournissent des informations, ces différentes contributions les guidant dans leurs propres décisions (Gainer et al. 2003; Rashid et al. 2006; Trevino 1999).

Les sondages posent des questions aux usagers pour les aider à comprendre l'importance de la CEL et tentent de comprendre pourquoi ils n'y contribuent pas (Dantzig 2002; Koh et al. 2007; Rothaermel et Sugiyama 2001; Wang et Fesenmaier 2003). Certaines de ces études incitent à « évaluer » le travail d'une autre personne, ou demandent à un public cible de fournir ses « opinions ». Elles incluent des emails qui implorent les usagers à contribuer à la CEL (Keen 1980). Les méthodologies dans ces études varient mais sont principalement centrées sur des sondages. Quelques autres méthodes consistent en expériences (Skretta 2007) et des expériences sur le terrain (Keen 1980). D'autres études expliquent les CEL en donnant un aperçu général (Skretta 2007), une histoire (Skretta 2007), une analyse anthropologique (Wilson et Peterson 2002), des explications de séries chronologiques (Wang et al. 2008), et des mesures de leur réussite (Preece 2001).

1.2.3 Bibliométrie

Cette étude prolonge ma recherche utilisant l'indice de Hirsch pour mesurer le statut de l'utilisateur et sa productivité dans l'affichage. L'indice de Hirsch, une mesure utilisée dans la bibliométrie, évalue à la fois la productivité et l'influence des auteurs. L'ancienne mesure d'influence d'un auteur a utilisé des sondages (Hirsch 2005). Dans le domaine de recherche des CEL on voit la recherche par sondage ciblant les propriétaires ou les membres de petites CEL (Law et al. 2008; Marett et al. 2009; Schaedel et al. 2010; Shen et al. 2008; Shen et al. 2009; Zheng et al. 2007).

La recherche par sondage, qui repose sur la capacité de l'enquêteur à évaluer les contributions aux CEL, est problématique à plusieurs égards. Premièrement, l'enquêteur a des opinions et des sentiments qui peuvent influencer les résultats du sondage. Deuxièmement, l'instrument de sondage

est inexact. Typiquement, une enquête repose sur des mesures correspondant aux points 5 ou 7 dans l'échelle de Likert. Cette mesure inexacte introduit des erreurs de la part de l'enquêteur (par exemple, répondre à une question par 5 ou 6) et du chercheur interprétant les données. Troisièmement, les personnes interrogées peuvent être un groupe auto-sélectionné, ce qui modifie les données. Dans un ensemble de données, les personnes qui répondent à des sondages peuvent représenter un groupe qui est en corrélation avec des personnes susceptibles de répondre à un sondage alors que ceux qui ne répondent pas ne sont jamais représentés. Par exemple, les « observateurs passifs » ont tendance à ne pas afficher ni répondre aux messages. Les « observateurs passifs » ne remplissent probablement jamais un questionnaire et n'apparaissent pas dans un sondage. Je propose d'utiliser une mesure bibliométrique qui considère l'influence en examinant quantitativement des réponses et opinions réelles (qui se lisent ou s'affichent sur un écran d'ordinateur).

L'indice de Hirsch a été proposé par Hirsch en 2005 et est défini de la manière suivante : « Un scientifique a un indice h si h de ses N_p articles ont chacun au moins h citations, et les autres ($N_p - h$) articles ont au plus h citations chacun. » Pour trouver l'indice h d'un auteur, on rassemble toutes les publications de cet auteur et les citations de ces publications. On classe ses publications dans une liste en ordre décroissant. Lorsque la i -ème publication (à partir du haut de la liste) a moins de i citations, on trouve l'indice h pour cet auteur.

L'indice Hirsch a été proposé comme un moyen de considérer les messages affichés d'une CEL dans le passé (Gomez et al. 2008; Laniado et al. 2011). Gomez et al. (2008) regardent les fils de discussion sur Slashdot, un site web populaire spécialisé dans l'information sur la technologie. Les chercheurs ont pris les messages et ont développé une « mesure pour évaluer le degré de controverse provoquée par un message » (Gomez et al. 2008). Ils le font en créant une arborescence à partir de chaque message initial. Les branches de l'arborescence sont des réponses au message initial et les réponses qui suivent. L'index h de Gomez et al. mesure le niveau d'imbrication maximale de l'arbre i où i « a au moins des commentaires égaux à $h > i$, ou en d'autres termes, $h + 1$ est le premier niveau d'imbrication i qui a moins de i commentaires » (Gomez et al., 2008).

L'article de Laniado et al. (2011) se penche sur les pages de discussion de Wikipedia en essayant de caractériser les discussions dans la section des commentaires. Laniado et al. (2011) fait cela en utilisant l'indice h pour mesurer la profondeur de la discussion sur un commentaire initial ainsi que la portée de la discussion. Laniado et al. (2011) construit des arborescences de discussion, comme le fait Gomez et al. (2008). Laniado et al. (2011) crée l'arborescence en utilisant les réponses qui deviennent les nœuds enfants pour l'entrée de discussion initiale. Pour Laniado et al. (2011) «

l'indice h de l'arborescence est alors le niveau maximal, pour laquelle le nombre correspondant de nœuds est supérieur ou égal à ce niveau » (Laniado et al., 2011). Ceci est similaire à la version de l'index h de Gomez et al. (2008). Une différence essentielle entre l'index h d'origine et leurs définitions de l'index h est que l'unité de mesure pour ces pages de discussion est le message initial qui mesure l'indice h au niveau des messages et non les auteurs de ces messages. Ils sont en mesure de trouver des messages initiaux controversés et porteurs, mais en utilisant leur indice h, ils ne peuvent pas identifier les contributeurs individuels responsables de ces messages.

J'utilise la mesure de l'indice h pour définir l'influence d'un contributeur à une CEL. J'interprète les citations dans la CEL comme des réponses aux messages envoyés. Mon utilisation de l'indice h correspond à la définition initiale de l'indice h dans la mesure où je tente de trouver les auteurs des messages porteurs. Ce qui m'intéresse n'est pas de trouver les messages porteurs, mais d'identifier les auteurs qui créent un contenu intéressant, qui est lu et commenté par un grand nombre d'autres usagers de la CEL. Je diffère de Gomez et al. (2008) et Laniado et al. (2011) dans la mesure où ces auteurs ont utilisé l'indice h pour mesurer l'influence dans les messages individuels, tandis que je garde l'indice h d'origine (Hirsch 2005) qui mesure l'influence d'un auteur. Je me différencie de Hirsch (2005) en ce que je mesure l'impact de l'auteur d'un message dans une CEL par des commentaires plutôt que par des citations.

L'autre avantage distinct que j'ai sur les études bibliométriques est que je peux savoir qui a consulté (ou au moins affiché sur son ordinateur) les messages d'autres usagers de la CEL. Donc, en plus de la réponse (citation), j'ai une deuxième mesure d'indice h qui prend en compte si les messages sont vus. Je définis ces critères comme l'indice hr (réponses) et l'indice hv (consultations). Ainsi, l'indice hr, défini comme un contributeur, a l'indice hr si hr démarrages de fil Np ont au moins hr commentaires chacun, et les autres démarrages de threads (Np - hr) ont au plus hr commentaires chacun. De la même manière, je définis l'indice hv quand un contributeur a l'indice hv si les hv démarrages de thread Np ont au moins hv commentaires chacun, et les autres démarrages de threads (Np - hv) ont au plus hv commentaires chacun.

1.2.4 Analyse du réseau Social

J'utilise également plusieurs mesures venant de l'Analyse de Réseau Social (ARS). J'utilise les mesures de centralité de degré, centralité de proximité, et centralité d'intermédiarité. La notion de centralité dans l'ARS se réfère au fait que, dans un réseau, il y a des nœuds qui sont au cœur du réseau, c'est-à-dire des individus qui sont au cœur du groupe. Si une personne est centrale au groupe, cela signifie

qu'un grand nombre de personnes est lié avec elle. Le degré de centralité est le simple décompte du nombre de connexions qu'on a dans le groupe (Freeman 1977). La personne théoriquement complètement connectée aura $n-1$ connexions (où n est le nombre de personnes dans le groupe) c'est-à-dire des liens avec tous les membres du groupe, sauf elle-même. Quand une personne est entièrement connectée la distance entre elle et le reste du groupe est 1 (Acedo et al. 2006; Eaton et al. 1999; Henry et al. 2007; Otto et al. 2002; Polites et al. 2008; Vidgen et al. 2007). Le degré de centralité est une mesure de combien de personnes on connaît dans un réseau.

La centralité de proximité mesure la distance entre les individus du groupe. La distance est mesurée par le nombre minimal de connexions pour accéder à cette autre personne. Donc, on doit d'abord trouver la distance entre un individu et tous les autres individus dans le groupe $n-1$. La centralité de proximité trouve alors la moyenne de ces distances (Polites et al. 2008; Vidgen et al. 2007). C'est une mesure de la façon dont on est proche du reste du réseau.

La centralité d'intermédiation mesure le rôle d'intermédiaire qu'on a entre deux personnes dans un réseau. Si on se trouve à une distance minimale entre deux personnes, on est « l'ami qui connaît un ami ». Donc, on est un intermédiaire crucial pour établir un lien entre des personnes qui veulent entrer en contact. L'intermédiation trouve les distances les plus courtes qui passent par une personne, puis additionne ces distances et divise par $(n-2)(n-1)$ (Vidgen et al. 2007). L'intermédiation est une mesure de l'importance d'un intermédiaire pour établir des liens entre les individus dans le réseau.

1.3. Contexte de la recherche

Quelques sections de cette thèse se concentrent sur le cas de BigUFans, un site web indépendant et autonome qui sert et constitue une CEL de supporters des équipes sportives d'une grande université aux États-Unis. Le propriétaire du site cherche à mieux mesurer la valeur des usagers en allant au delà du simple comptage des messages. Il s'est rendu compte que le maintien d'un groupe d'usagers de base réguliers est la clé du succès d'une CEL. Les autres membres attendent généralement que ces membres clé se manifestent avant d'afficher leurs réponses. Cependant, ce propriétaire de CEL s'est également rendu compte que la quantité de messages n'était pas le seul critère pour identifier le groupe d'usagers de base. Au lieu de cela, dans une CEL comme BigUFans, la qualité des messages et la réputation du participant influencent la façon dont d'autres membres s'engagent sur un sujet donné. La CEL BigUFans est représentative de nombreux sites semblables, qui veulent attirer les

publicitaires vers un public spécifique constitué de supporters dévoués. C'est par la publicité que le site a, depuis près de 15 ans, obtenu suffisamment de revenus pour se maintenir en ligne et améliorer sa capacité de serveur, qui, à son tour, a rendu le site plus facile à utiliser. Toutefois, c'est le dynamisme de la communauté qui lui donne sa pertinence. Une meilleure compréhension de quels usagers sont les moteurs de ce dynamisme est le but de ma recherche.

BigUFans.com est un site de CEL pour l'institution Big Université. Big U. est une grande université de recherche de niveau I dans le sud-est des États-Unis. Big U. est une université publique fondée il y a 150 ans avec plus de 30.000 étudiants. Cette CEL est orientée vers les équipes sportives de Big U. où le football américain est le sport principal. Big U. participe à la Division I de la NCAA (National Collegiate Athletic Association) dans de nombreuses disciplines sportives, y compris les sports clés que sont le football américain, le basketball et le baseball.

Actuellement BigUFans.com a quatre concurrents majeurs. BigUFans.com a été créé en 1999 et est la plus grande CEL d'accès libre parmi cinq CEL concurrentes. Puisqu'elle produit des revenus par la vente de publicité, l'activité sur BigUFans.com représente une source importante de revenus. Les quatre sites concurrents ont des parties de leur CEL qui sont payantes. Toute personne peut anonymement regarder le contenu de BigUFans.com mais pour afficher des messages ou participer à des forums, un usager doit s'inscrire.

Il y a 6 administrateurs et 17 modérateurs qui travaillent pour assurer la pertinence des sujets sur le forum de BigUFans.com et pour éliminer le spam et des commentaires incendiaires. Les membres inscrits peuvent devenir modérateurs s'ils deviennent plus actifs, mais seulement sur invitation du propriétaire du site. En raison du travail du personnel, l'un des principaux avantages de BigUFans.com sur ses concurrents est que la communauté se perçoit comme bénéficiant de messages plus positifs, exempts de publicité, de spam, de propos incendiaires, et de discussions hors-sujet.

La majorité de l'activité sur BigUFans.com vient de la discussion centrée sur le football américain. Au cours de la saison 2011-2012, l'activité mensuelle a vu une moyenne de 2 900 000 pages consultées, 1075 nouveaux messages, 167.000 visiteurs, et 89 sujets abordés. L'activité quotidienne inclut 1017 usagers inscrits et 5 nouveaux usagers. Une visite typique se compose de la consultation de 5, 2 pages. D'un point de vue démographique, les inscrits sont typiquement des hommes (92%), diplômés universitaires (72%) et ayant des revenus supérieurs (71%). Pendant l'intersaison avant la saison 2011-2012, il y a eu 2 millions de pages consultées par mois. Actuellement, BigUFans.com accorde automatiquement des statuts de niveau différent pour les usagers qui affichent un certain nombre de messages, quel que soit le nombre d'usagers qui les lisent. BigUFans donne les statuts d'usager

suivants (avec un nombre minimal de messages requis) : Équipe Scout (0 messages minimum requis), Troisième Équipe (13 messages), Deuxième Équipe (50 messages), Première Équipe (250 messages), All-Conférence (1000 messages), All-American (2000 messages), et Hall of Fame (5000 messages). En ce moment, il y a 20 470 membres dans l'Équipe Scout, 1363 membres dans la Troisième Équipe, 1086 membres dans la Deuxième Équipe, 590 membres dans la Première Équipe, 179 membres dans la catégorie All-conférence, 121 membres dans la catégorie All-American, et 47 membres dans la catégorie Hall of Fame. L'appartenance à l'Équipe Scout est essentiellement accordée à toute personne qui s'inscrit pour le service. Il y a en outre 6 administrateurs et 17 modérateurs. Ces fonctions sont attribuées aux usagers qui sont des membres établis et responsables (sans spam) et qui acquièrent ces responsabilités sur invitation.

1.4. Présentations à des conférences

Les communications suivantes ont été présentées lors de conférences, soumises à des journaux académiques ou sont en cours de soumission. Ma recherche a donné lieu à cinq présentations à des conférences.

1.4.1 Étude de l'impact des événements hors ligne sur le développement du groupe dans une communauté sportive en ligne

Cette communication a été présentée en 2011 à la Southern Association for Information Systems Conference à Atlanta. Elle examine l'état de la recherche sur les CEL et a trouvé que les communautés qui réussissent ont des membres qui créent et partagent des liens et identités communs. L'objectif de cette recherche en cours est de mieux comprendre l'impact que des expériences partagées ont sur les activités des usagers. La recherche a utilisé PLS dans une analyse préliminaire des relations entre les pages consultées de la CEL BigUfans.com et les événements réels qui affectent les membres de cette communauté. Les résultats préliminaires suggèrent que notre modèle représente 60, 4% de la variance de pages consultées sur le site et 27, 2% des contributions (messages sur forums de discussion) sur le site.

1.4.2 Influence de l'utilisateur d'une communauté en ligne: étude avec User Status

Cette communication a été présentée en 2012 à la Southern Association for Information Systems Conference à Atlanta. Cette recherche en cours essaie de comprendre la relation entre l'influence et le statut des usagers dans une CEL. Un enjeu majeur est de savoir comment continuer à prospérer et se maintenir dans le long terme. Souvent, les CEL gagnent en popularité, et puis la perdent en raison de la stagnation du nombre de nouveaux usagers et de la baisse d'intérêt des usagers inscrits. Cette recherche tente de comprendre le statut des usagers et quelle influence les « super usagers » ont sur la vitalité d'une CEL. Nous mesurons l'influence des usagers en utilisant un pseudo indice de Hirsh qui mesure l'intérêt pour les messages affichés et en évaluant le statut des usagers de la CEL.

1.4.3 Quels facteurs déterminent la contribution des usagers dans une communauté en ligne? Etude sur l'influence des contributeurs et sur le statut des usagers

Cette communication a été présentée en 2012 lors de la America's Conference on Information Systems à Seattle. Les CEL dépendent des intérêts communs et des interactions entre usagers facilitées par la technologie. Les CEL qui prospèrent trouvent des moyens d'encourager ces interactions pour élargir leur base. Beaucoup de CEL ont des usagers influents qui aident à agrandir la communauté par leur présence et leurs contributions. Cependant, le processus d'identification de ces usagers n'est pas trivial. Cette étude propose une nouvelle méthode pour identifier ces usagers à grand impact par la création d'indices de Hirsch modifiés. Cette méthode améliore les méthodes actuelles qui se contentent soit de compter les contributions soit de sonder d'autres usagers. Nous validons ces nouvelles mesures du statut des usagers et analysons ensuite les mesures en les corrélant avec les messages affichés, les démarrages de threads, et les consultations et réponses aux démarrages des fils de discussion pour une CEL basée sur un intérêt commun.

1.4.4 Influence d'une communauté en ligne: étude utilisant l'analyse métrique de réseau social de Hirsch

Cette communication a été présentée en 2013 lors de la America's Conference on Information Systems à Chicago. Cette étude porte sur les petites et moyennes CEL et tente de trouver des moyens de mesurer l'impact des contributions d'usagers, dont les CEL sont tributaires pour prospérer. Pour les parties prenantes, un enjeu important est de mesurer le dynamisme des CEL en identifiant les usagers qui ont le plus d'influence et qui créent le plus d'activité, augmentant ainsi le nombre de consultations. Afin d'évaluer ces usagers performants, nous étendons la recherche

précédente par une mesure en deux parties. Tout d'abord, en utilisant les paramètres de Hirsch pour mesurer la productivité et l'impact des contributions d'usagers et ensuite, en exécutant une analyse de réseaux sociaux qui considère les usagers dont le rôle est central dans le réseau. Cette étude, centrée sur un site de supporters d'une équipe sportive universitaire, a pour objectif de mesurer l'influence de ses usagers et a trouvé une certaine corrélation entre les mesures de Hirsch et les mesures de centralité.

1.4.5 Mesurer la contribution d'un nouveau membre dans une communauté en ligne: une approche par l'attachement préférentiel

Cette communication a été soumise en 2015 à la International Conference on Information Systems (ICIS) à Fort Worth. La CEL est une forme populaire de transfert de connaissances spécialisées, où des usagers géographiquement dispersés peuvent former une communauté en partageant des idées, envoyant des messages et forgeant des amitiés en ligne. Les CEL ont tendance à avoir un cycle de vie limité, où leur apparition et leur croissance initiale sont suivies par une phase de stagnation où les usagers cessent d'afficher des commentaires, ce qui mène la communauté à mourir par manque d'activité. Une façon dont une CEL peut rester dynamique est d'encourager ses nouveaux usagers à contribuer. Dans cette recherche, nous évaluons les nouveaux usagers d'une CEL pour leur attachement préférentiel, c'est-à-dire leur tendance à préférer afficher leurs messages sur un fil de discussion dont les auteurs sont connus et respectés dans la CEL. Nous examinons les niveaux d'affichage des usagers existants pour mesurer l'impact des usagers déjà établis dans la CEL et comparer le comportement des nouveaux usagers à celui de ces usagers chevronnés. Nous rendons notre méthode opérationnelle en utilisant un site de supporters d'une équipe sportive universitaire.

1.4.6 Mesurer l'impact des contribution des membres d'une communauté en ligne: une nouvelle approche inspirée par l'indice de Hirsch

Cet article est le premier de ce projet de recherche à avoir été soumis à une revue académique: il s'agit d'un numéro spécial de ISR en 2015. Il poursuit les travaux présentés à la conférence SAIS et AMCIS en 2012 (voir sections 4.2 et 4.3). Une CEL est une forme très répandue de transfert de connaissances spécialisées, où les usagers géographiquement dispersés peuvent constituer une communauté en partageant des idées, envoyant et affichant des messages, débattant de sujets et nouant des amitiés en ligne. Un des problèmes de ces CEL est qu'elles ont tendance à avoir un cycle

de vie limité, où leur apparition et leur croissance initiale sont suivies d'une phase de stagnation où les usagers cessent d'afficher des commentaires, ce qui mène la communauté à mourir par manque d'activité. Essayer de prolonger la phase de croissance et trouver des moyens de maintenir son dynamisme en encourageant les contributions est une préoccupation fondamentale pour tout administrateur de CEL.

Dans cette recherche, je propose d'utiliser une mesure bibliométrique non-invasive utilisant la méthodologie de l'indice de Hirsch pour identifier des contributeurs performants. Je défends l'idée que ces usagers, dont l'identité est bien établie et vérifiée grâce aux réponses et consultations dont bénéficient leur fil de discussion contribuent substantiellement aux CEL, ce que mesure l'indice de Hirsch. J'ai testé les données sur un site de supporters d'une équipe sportive universitaire. Je les ai aussi analysées pour identifier les contributeurs dont le rôle est crucial pour la CEL. Enfin, j'ai démontré comment utiliser cette nouvelle mesure pour une CEL réelle.

1.5. Analyse et répercussions

Ces recherches apportent plusieurs contributions et présentent aussi des limites.

1.5.1 Contributions

Le critère de statut de l'utilisateur a cours au moment où un usager commence à afficher un certain nombre de messages dans une CEL. Bien que ceci soit excellent pour encourager les usagers à revenir et afficher des messages fréquents, je considère ces statuts artificiellement gonflés. Puisque cette mesure est un produit des affichages considérés depuis le démarrage d'une CEL, je pense qu'une approche plus démocratique, qui tient compte du comportement des autres usagers, serait une bien meilleure mesure de son influence. Un ensemble de données, telles que les mesures SNA, considère le comportement des usagers. Un ensemble de données, telles que les indices h , mesurent le comportement des usagers dans leurs consultations et leurs réponses à d'autres usagers.

En outre, je trouve qu'une mesure globale est une mesure bien meilleure en ce qu'elle peut facilement révéler un bref horodatage des données pour trouver ce qui est pertinent dans une période récemment écoulée, qu'elle soit de quelques semaines ou de quelques mois. Je pense aussi que l'utilisation de la mesure globale me permet de trouver différentes catégories d'utilisateurs. J'ai commencé à en identifier certains, tels les « agitateurs », « observateurs passifs », ou « fournisseurs d'information ». Je sens que des recherches futures pourront utiliser ces catégories pour permettre

aux propriétaires de CEL de cibler différents usagers pour de potentiels changements de comportement : essayer d'amener l'« observateur passif » à contribuer davantage ; viser l'« agitateur » qui tient des propos incendiaires par d'éventuelles mesures disciplinaires, comme le renvoi; ou encore encourager le « fournisseur de l'information » à offrir des messages susceptibles de générer plus de discussions.

1.5.2 Répercussions pour les professionnels

Je pense que cette mesure donnera aux propriétaires de CEL de meilleurs outils d'évaluation des usagers. La méthode actuelle, basée uniquement sur la quantité de messages ou la création de contenu, ne donne pas une mesure adéquate de l'impact. Pour les professionnels, l'intérêt devrait être la façon dont certains messages ont plus d'influence que d'autres. Cela signifie qu'une mesure comme celle que je propose devrait être utilisée pour encourager les usagers qui participent.

Une autre conséquence est qu'utiliser cet instrument de mesure pour faire émerger différents types d'usagers, tels l'« agitateur », l'« observateur passif », ou le « fournisseur d'information ». Cela permet aux professionnels d'identifier ces différents types d'usagers, de choisir des formes d'encouragements en fonction de ces types et de se donner les moyens d'améliorer le dynamisme des CEL.

Cela permet également une analyse globale des CEL avec la possibilité d'utiliser des données d'archives. Par rapport aux méthodes de sondage, qui peuvent biaiser les résultats en raison des contraintes qu'ils génèrent dans les réponses, utiliser des mesures non invasives facilite la collecte des données.

1.5.3 Limites

Bien que les mesures identifiées proposent une amélioration par rapport aux méthodes antérieures, elles souffrent de quelques limitations. D'abord l'ensemble de données utilisé n'est pas généralisable à d'autres CEL. Les données présentes ont été collectées dans une CEL de supporters d'équipe sportive universitaire. La généralisation non seulement à d'autres CEL de supporters d'équipes universitaires ou à d'autres CEL de supporters sportifs, mais plus généralement à d'autres types de CEL pose problème. Une CEL de supporters sportifs a aussi l'avantage d'avoir des sujets de conversation renouvelés lors de chaque saison ainsi que de match en match. Pour un autre type de

CEL, par exemple, une CEL de jeux informatiques, une fois que les membres de la communauté ont fini de jouer et que les conversations sur ces jeux sont épuisées, les sujets de conversation manquent pour continuer la CEL. Pour une CEL de supporters sportifs la situation est différente car de nombreux sujets de conversation persistent de match en match. De nouveaux joueurs peuvent être recrutés, de nouvelles blessures peuvent les affecter, et les stratégies peuvent être discutées d'un match à un autre. C'est pourquoi la généralisation de l'ensemble de données peut être remise en question. En outre, j'ai utilisé une période de six mois et ai été en mesure d'obtenir un très grand ensemble de données. Le type de CEL vers laquelle généraliser peut être différent si cette CEL n'a pas ce type d'activité hebdomadaire.

Une autre limitation peut venir de l'utilisation des réponses et consultations. Comme les indices h_r et h_v ont été dérivés de l'indice h , les réponses et les consultations pour h_r et h_v , respectivement, ont remplacé les citations comptées par l'indice h . On peut être d'avis que les réponses à un fil de discussion de CEL n'ont rien de comparable au travail qui entre dans une citation. Qui plus est, pour un membre d'une CEL les consultations sont encore plus faciles que les réponses. On peut en outre faire valoir que l'indice h_v pourrait être moins important, car les consultations ne signifient pas nécessairement que l'on a « lu » le message, mais seulement qu'on a cliqué et affiché le message sur l'écran. Une autre question qui se pose concernant la plus grande facilité des consultations par rapport aux réponses est que, avec des consultations plus faciles, l'indice h_v est considérablement plus élevé que l'indice h . Ainsi, alors que l'indice h_v peut avoir moins de poids en raison de la facilité des consultations, l'indice h_v est susceptible d'être plus élevé, ce qui rend la comparaison des deux plus difficile.

Une troisième série de limitations vient de l'utilisation des méthodes de l'indice h dans ces mesures. L'indice h a été critiqué sur plusieurs points. Les principaux problèmes évoqués sont le fait que l'indice h est plus difficile à obtenir par des chercheurs plus jeunes, l'indice h ne mesurant pas la recherche récente et ne tenant pas compte de la portée d'une citation. Les deux premières critiques de l'indice h sont intimement liées. Un fait sur l'indice h est qu'il ne peut jamais descendre. Cela est dû au fait qu'une fois qu'on obtient une citation, cette citation ne disparaît jamais. Donc, si un chercheur obtient un indice h de valeur 1, ce chiffre ne pourra jamais revenir à zéro. Cela est lié au fait qu'un jeune chercheur a plus de difficulté à obtenir un indice qu'un chercheur plus chevronné, car des publications récentes obtiennent difficilement un indice h . L'indice h est lié aux citations et publications récentes; il faut du temps pour obtenir ces citations. On ne peut pas prédire la tendance des citations dans l'avenir, ce qui explique qu'un jeune chercheur ait plus de difficulté à d'obtenir un indice h .

Le dernier problème avec l'indice h est que les documents qui recueillent un grand nombre de citations, les « ouvrages à succès isolés », ont tendance à être négligés par l'indice h. L'indice h est une mesure de la productivité de la publication et de l'impact de ces publications par des citations. Si un chercheur a publié deux articles dans sa carrière, mais ces deux articles ont reçu des milliers de citations, l'indice h pour ce chercheur ne dépassera pas 2. En raison de la définition de l'indice h, qui compte le nombre de publications, l'indice h maximal que l'on peut obtenir est limité par le nombre de publications que le chercheur a publiées. Dans cette même veine, les indices h_r et h_v sont limités au nombre de threads qu'un membre d'une CEL a démarrés.

1.6. Tableau de synthèse des recherches

Le tableau suivant montre les publications qui ont été présentés et sont actuellement en cours. Un total de cinq communications a été présenté à des conférences, une communication a été soumise à une conférence et un article a été soumis à un journal académique. Les recherches futures devront continuer dans ce courant de recherche. Le tableau ci-dessous résume la recherche.

Table 1-1. Tableau des Recherches

Titre	Objectif	Contribution	Théorie	Méthodologie	période de collecte de données	Résultats	Ma Contribution	Futur	Conférences
Étude de l'impact d'événements hors ligne sur le développement du groupe dans CEL de supporters sportifs (Ch. 3)	Trouver le lien entre des événements réels et les activités d'une CEL	Activité d'une communauté virtuelle dans ses liens avec la vie réelle	Lien commun / Identité Commune	Modèle conceptuel utilisant programme d'amorçage et PLS	trois mois (août 1, 2009 à janvier 31, 2010)	Incidents de la vie réelle ont un effet important sur les contributions à une CEL	L'analyse des données, la majorité de l'écriture, la présentation à la conférence	Collecte de données et soumission à un journal académique	SAIS 2011
Influence de l'utilisateur d'une CEL: étude avec User Status (Ch. 4)	Le statut de l'utilisateur dans une CEL correspond-il réellement à celui d'un contributeur performant ?	Validité des indices hr et hv pour mesurer la contribution à une CEL	Bibliométrie	Utiliser les indices hr and hv pour mesurer l'influence des messages et vérifier si le statut de l'utilisateur lui correspond	Pas de données prises, proposition de recherche seulement	Description de la recherche	la majorité de l'écriture, la présentation à la conférence	Précurseur à communication à AMCIS 2012 (Ch. 5) & soumission à journal professionnel (Ch. 8)	SAIS 2012
Quels facteurs déterminent la contribution des usagers dans une CEL? Etude sur l'influence des contributeurs et le statut des usagers (Ch. 5)	Trouver si hr et hv prédisent mieux les contributeurs performants	Validation empirique des indices hr et hv	Bibliométrie	Analyse de corrélation entre hr, hv, messages, démarrages de threads, consultations, et réponses sur période de 1 et 2 ans.	Trois ans, 2009 à 2011	Statut d'utilisateur, hr, et hv prédit 1 an de consultations et réponses dans CEL mais hr et hv sont mieux que statut d'utilisateur pour prédire sur 2 ans	Auteur principal, la majorité de l'écriture, l'analyse de données, la rédaction du programme Java, présentation à la conférence	Soumission à un journal professionnel (Ch. 8)	AMCIS 2012
Influence d'une communauté en ligne: étude utilisant l'analyse métrique de réseau social de	Vérifier la corrélation entre la centralité du réseau social et l'index h pour contribution d'utilisateur	Ajout de la recherche en Analyse de Réseaux Sociaux (ARS)	Bibliométrie et Analyse de Réseaux Sociaux (ARS)	Analyse de corrélation entre mesures de ARS et indices h	Six mois, (août 2012 à février 2013)	Démonstration de hr et hv pour la centralité de degré et d'intermédiation	La majorité de l'écriture, l'analyse des données, l'écriture d'un	Collecte de données pour soumission à un journal professionnel	AMCIS 2013

Hirsch (Ch. 6)	dans une LEC					mais pas la centralité de proximité	programme Java, présentation à la conférence		
Mesurer la contribution d'un nouveau membre dans une CEL : approche par l'attachement préférentiel (Ch. 7)	Expliquer la durabilité et le dynamisme d'une CEL par le comportement de nouveaux usagers et les motiver à augmenter leur niveau de contribution	Utilisation de l'attachement préférentiel pour expliquer comportement de nouveaux usagers dans une CEL	Attachement préférentiel, analyse de réseau	Analyse de comportement de nouveaux usagers pour expliquer les hypothèses	Un an pour les nouveaux utilisateurs (2014) pour les données d'anciens recueillies pendant 15 ans (1999-2014)	Aide aux nouveaux usagers pour démarrer de nouveaux <i>threads</i> , aux usagers chevronnés pour répondre plus, et répondre plus aux autres usagers chevronnés	Auteur principal, la majorité de l'écriture, l'analyse de données, la rédaction du programme Java	Attente de décision sur soumission à une conférence	ICIS 2015
Mesurer la contribution d'un nouveau membre dans une CEL: approche par l'attachement préférentiel (Ch. 8)	Trouver un meilleur moyen de mesurer la contribution des membres d'une CEL	Expliquer les contributions des usagers avec hr, hv, et SNA et une mesure globale « somme de rang »	Bibliométrie, ARS	Calcul des mesures et analyse des résultats	Six mois (octobre 23 2012 à avril 21, 2013)	« Somme de rang » semble mesurer les contributeurs performants. Corrélation entre intermédiarité et hr, hv, mais proximité reste sans incidence sur ces mesures	Auteur principal, la majorité de l'écriture, l'analyse de données, la rédaction du programme Java, persentation à la conférence	Soumission à un journal professionnel	En cours

1.7. Conclusion

J'ai commencé cette recherche pour trouver de nouvelles méthodes qui permettent aux propriétaires de CEL de trouver des usagers qui contribuent à la vitalité de leur CEL. Les problèmes majeurs des CEL sont leur durabilité et leur vulnérabilité face à la potentielle perte d'information pertinente qui leur fait trop facilement rejoindre le cimetière des CEL défuntes. J'ai proposé une mesure globale utilisant des mesures SNA et bibliométriques qui dépendent du comportement des membres des CEL. Ces mesures se présentent comme plus descriptives et révélatrices que celles—actuellement en cours—qui considèrent seulement le statut de l'utilisateur ou les fréquences d'affichage.

J'ai utilisé les données de BigU, une communauté en ligne pour les supporters d'une équipe sportive universitaire. Les données ont été collectées sur une période de six mois et ont montré la diversité des principaux contributeurs à la CEL. Je suis en mesure d'identifier plusieurs catégories d'utilisateurs qui peuvent permettre à un propriétaire de CEL d'identifier les usagers les plus productifs.

Chapitre 2

General Introduction (EN)

The online community (OC) is a popular form of specialized knowledge transfer, where geographically dispersed users can form a community by sharing ideas, send and post messages, debate topics, and forge online friendships. One of the problems with OC's is that they tend to have a life cycle, where there is the birth and growth of the OC but then there is a stagnant stage where users stop posting to the OC and the community eventually dies due to inactivity. Trying to extend the vibrant growth stage of an OC is a relevant topic for any administrator of an OC. One way that an OC can stay vibrant is to encourage contributions.

This research stream will look at how OC's can keep their vibrancy for a longer period of time, by looking at various aspects of OC's such as measures of user contribution and how new users in an OC behave. I propose to use different measures to evaluate users contributions to an OC. One of these measures is a non-invasive bibliometric measure using the Hirsch-index methodology as a way to identify high-level contributors. Another stream of this research will look at how new users behave and how this might be explained by preferential attachment.

2.1. Introduction

Online communities (OC) continue to grow in popularity and sophistication as a means for geographically dispersed members that have similar interests to communicate and share ideas and thoughts and multimedia. These OC's offer a way for people with common interests to actively participate in sharing of information, debate relevant (or even irrelevant topics), forge online connections, and build a sense of being part of a vibrant OC. OC's are often used as a source of knowledge and a repository of information from which members draw on their real-life experiences. But in order for these communities to exist, there needs to be active participation by the members. The contributions by members need to also be relevant, correct, and up-to-date.

An online community is defined as an "aggregation of individuals or business partners who interact around a shared interest, where interaction is at least partially supported and/or mediated by technology and guided by some protocols or norms" (Porter 2004). Sometimes an OC is referred to as a virtual community or an Internet community. These OC's can be gathering places for people with similar interests where they can share information, talk about experiences, or just meet up with friends to chat.

While OC's pop up all the time, I have seen many OC's that have died off. OC's that have become irrelevant are often can be found in special community groups in Yahoo or Google. These groups may be familiar, as they tend to be forums where questions are or were posed and once sparked lively debate and discussion, but have either closed or become inactive due to lack of participation or relevance. Other now less popular OC's such as Second Life and Myspace are categorized as a social networking site but have OC community characteristics and can have OC components to them. These social networking sites have the ability to post and share information between users, which is a common characteristic of an OC. Vibrant and thriving OC's such as Tripadvisor, Flyertalk, and Rotten Tomatoes also exist. In each of these sites, there are ways to report reviews of products of interest, i.e. hotels for Tripadvisor or movies for Rotten Tomatoes; users can give feedback to the reviews, whether it was a helpful review or not.

Sometimes an OC is tied to an entity that in itself can become popular but over time diminish in popularity. There are many OC's dedicated to the gaming or sports community where the product or franchise may have been popular at one time but the product may not have had updates or the team may have moved (e.g. Red Dead Redemption or the Atlanta Thrashers). There are also many products that have been kept updated and the OC's tied to these products tends to live on (e.g. Halo, World of Warcraft).

While there may be a life-cycle effect for an OC, the owner's goal is to keep the community vibrant and to engage members. To engage members, the community must have topics and discussions that are relevant and timely. Many OC's need to earn revenue via advertising and the advertising dollars come in because of a mixture of views, click-throughs on ads, and purchases through ads. One such site is, BigU, which is an online community for a university sports team that will be introduced in more detail later.

For this research stream, I have an overall objective of finding ways for keeping an OC up and running. To do this, the OC must maintain revenue to cover the expenses of hosting, creating content, and maintaining the OC. To keep revenue up, the OC must maintain a good advertising base, which means, at the very least, many people must see the advertisements. However, to attract people to the OC, content must be stimulating and timely. Therefore, the targeted research question for this project is how does one keep the vibrancy of an online community?

For the purpose of this research, vibrancy refers to the interactivity in the OC. Vibrancy is a measure of the postings and subsequent views and replies to these postings in the OC. A vibrant OC then is an OC that has frequent posting submissions and much interaction with those posts. When an OC loses vibrancy the OC interaction become stagnant, postings become scarce, and view and reply activities on postings also cease. To keep the vibrancy, an OC needs to attract many people and that means user-created content must be interesting.

One problem area I found is a lack of ways to identify users that contribute relevant content for the community. There is the simple 'posting number' that many OC's use today to demonstrate how much a user has contributed to an OC. Unfortunately, post counts are simply a measure of productivity. Impact, influence, and whether other users are reading their contributions has no bearing in a pure count of posts. Therefore, the current research is attempting to find better ways to measure the influence that contributors have within the online community.

I propose an aggregate measure of using social network analysis and bibliometrics (Takeda 2010) to measure the influence and productivity of a user, and run data from an online community. In this research I find different types of users that were identified by this aggregate measure as opposed to only looking at the 'posting number' measure.

2.2. Background

I first look at OC's, focusing especially on OC health and how vibrancy of an OC is attained. I then look at two main areas where I derive the measures, bibliometrics and social network analysis.

2.2.1 Online Communities

This research is concerned with the health of an OC. I am also interested in how to maintain this health and vibrancy of the OC. OC's require interactions, contributions, and visits from the OC community. Unlike a e-commerce website selling a tangible product, an OC requires major interaction from the customers (whom are the community members) for its revenue stream, but the OC can also have a customer come back as a repeat customer over and over because the content can change so quickly. For example, in the case study OC, many users log in multiple times during the day because new content and news is posted throughout the day. As long as the OC content continues to add value to the customer, the OC owners can reap the benefits of providing this service.

There are many studies conducted on OC's (Harper et al. 2007; Iriberry and Leroy 2009; Ren and Kraut 2007; Sassenberg and Postmes 2002). OC's are usually formed by some topic or interest area. The users join the OC in order to get more information or share information about their common interest. Research in OC's have targeted various different types of OC's. For example there is research that has targeted sites on topics such as movies/entertainment (Keen 1980, Wang and Fesenmaier 2003), travel (Wang and Fesenmaier 2004), business (Wang and Fesenmaier 2003; Wang et al. 2008), health (Wang and Fesenmaier 2003), and computers/science (Wang and Fesenmaier 2003; Zhang et al. 2007). In the IS field there are studies that looked at knowledge management systems (Kankanhalli et al. 2005), as well as studies that looked at software development, addressing OC's that are concerned with open source software development (von Krogh et al. 2003; von Krogh et al. 2012), as well as packaged software development (Holström and Henfridsson 2006).

The current research on OC participation has evolved from the peer-to-peer (P2P) file sharing research. The P2P research tended to look at the contributor vs. downloader mix and how the community was riding on the work by the contributors (Golle et al. 2001). More recent research attempts to define some form of evaluation system to rate contributors and the impact of their contribution. For example, users that read the contributions might be solicited to provide feedback. The systems are typically set up to give credit to the contributor as well as the evaluator in the form of some sort of compensation (contribution points, virtual currency, etc.).

Whether an OC becomes and remains successful is heavily dependent on the participation by community members. The OC is only as good as the content created by the community members (Assmann et al. 2009; Cothrel et al. 1999). OC's tend to have a lifecycle where, in the beginning they are vibrant and relevant, a middle where they maintain vibrancy, and then an end where they fall out with the community and there are minimal posts and low interaction over a long period of time

(Iriberry et al. 2009). OC's that die out tend to have irrelevant or outdated contributions and members find the content no longer interesting and stop visiting the OC. Stagnation events lessen new contributions and, thus, lead to a vicious cycle of stagnant behavior by the community members. Even if there are many visitors to the OC, visits alone do not guarantee vibrancy since community interaction is not static (Mousavidin et al. 2009). There are studies of OC's addressing the downfall of OC's (Gainer et al. 2003; Keen 1980; Koh et al. 2007; Rothaermel and Sugiyama 2001; Trevino 1999). OC's may appeal to visitors so long as new and interesting material is available. However, as communities decrease frequency of new contributions, there is less incentive for current members to visit the community on a regular basis. Once an OC loses momentum, trying to win back prior visitors becomes much more difficult. Many studies have tried to address this problem by trying to identify why the community becomes stagnant and identifying ways to improve the contributions to an OC (Gainer et al. 2003; Preece 2001; Ren et al. 2007; Rothaermel and Sugiyama 2001; Trevino 1999). Therefore, a need exists in practice to ensure the vibrancy of the OC is not lost. The challenge for OC owners is to try extending the initial growth or middle vibrancy period as long as possible to maintain the health of the OC. While many papers have been written about large OC's such as Myspace, Facebook, and ebay I have found that research looking into the health of smaller OC's is lacking (Clavio 2008; Shen et al. 2006).

There is also research concerned with the identification of users that contribute. The 'user status' is a measure of the community users used by many OC's that is derived from the number of postings by a user. Typically the 'user status' is a rank for a user that is given after the user has contributed a certain amount of posts or been an active member for a particular length of time. For example, in a sports related OC a user might gain 'bench player', 'rookie', 'regular', 'all-star' status by posting 5, 10, 100, and 1000 times, respectively. User status in OC's has been an area of research in the past. This area relies on social capital (Ganley et al. 2006; Law et al. 2008), social presence (Shen et al. 2009; Shen et al. 2006), social identification (Shen et al. 2008; Zheng et al. 2007), and social influence theory (Assmann et al. 2009; Cheng et al. 2006). Some of these studies try to make the connection between high status users and the vibrancy and thus the health and success of the OC (Assmann et al. 2009; Cheng et al. 2006).

I argue that 'user status' is only a count of production and by no means is it a measure of how the postings are received by the community members. One can post thousands of posts but all of those posts may be read by no one in the OC, thus not really contributing anything to the health and vibrancy of the OC.

Some studies have evaluated why there is a loss in community member activity and how to help stagnant communities become more active (Harper et al. 2007). Several studies have shown that lurkers exist in all OC's and that this form of 'social loafing' is detrimental to communities (Shiue et al. 2010). Some studies identify 'loafing' or 'lurking' as a contributor to the demise of an OC (Dantzig 2002; Gainer et al. 2003; Keen 1980; Rothaermel and Sugiyama 2001). Loafing is "the tendency for individuals to expend less effort when working collectively than when working individually" (Shiue et al. 2010 p. 769). Lurking is the activity of reading "messages on an Internet discussion ... without contributing" (Merriam-Webster Online Dictionary 2010). Member non-participation may indicate dissatisfaction and the lack of a compelling reason for the lurker to become active. Lurkers may be less satisfied than active posters or do not want to add confusion to already-busy message boards (Preece 2004).

There has been some research trying to change the 'lurkers' into 'contributors'. These studies have tried to identify the motivation for contributors to contribute. The reasons for contributors to contribute have been identified as being a way to build up one's reputation online (Kankanhalli et al. 2005; Ma and Agarwal 2014; Ren et al. 2012; Wasko and Faraj 2005) and/or contributors contribute knowledge just because it is satisfying to help others (Kankanhalli et al. 2005; Wasko and Faraj 2005). Other studies have looked specifically at the feedback mechanism to allow the contributors to know when they are gaining reputation in the OC (Burke et al. 2009; Cheshire and Antin 2008; Jung et al. 2010).

2.2.2. Invasive Methods

Two types of methods frequently occur in the literature: survey studies and evaluation studies. Both types require some form of user input. Feedback studies have some inherent feedback mechanism built into the OC. Typically these systems include functionality allowing users to 'rate' the work of another user or the target unit of measure of the website. For example, some OC's have the ability for a user to rate a contribution by another user. If there was a question posed by a user and several people answered the question to varying degrees of detail, a user would be able to rate each of the answers given. While it is possible to have a multi-level rating system, typically these types of sites would only allow a 'Yes/No' type of Boolean rating system (Gainer et al. 2003; Keen 1980; Trevino 1999). An example of a study conducted on a unit of measure was one study that targeted a movie review website (Keen 1980). Feedback was used by other members of the website to see a composite rating and user comments. These feedback systems typically allowed users to see others ratings to allow users to gain information to make their own decisions about different contributions (Gainer et al. 2003; Rashid et al. 2006; Trevino 1999).

Survey studies ask users questions to help users understand the importance of the community and try to understand why they were not contributing to the OC (Dantzig 2002; Koh et al. 2007; Rothaermel and Sugiyama 2001; Wang and Fesenmaier 2003). Some of these studies required some form of incentive to 'rate' another's work, or would solicit 'ratings' from a target audience. These included emails that were sent to the users imploring them to contribute to the OC (Keen 1980). Methodologies in these studies were varied but centered mostly on surveys. Some other methods seen were field experiments (Keen 1980) and experiments (Rashid et al. 2006). Other studies explained OC's by giving an overview (Skretta 2007), history (Skretta 2007), anthropology (Wilson and Peterson 2002), time series explanation (Wang et al. 2008), and success measures (Preece 2001).

2.2.3. Bibliometrics

This paper extends my research of using the Hirsch index to measure the user status and posting productivity. The Hirsch index has been a measure used in bibliometrics that measures both productivity and influence of the authors. The old measure of author influence has been using survey research (Hirsch 2005). In the OC research area one sees survey research targeting the owners or community members of small OC's (Law et al. 2008; Marett et al. 2009; Schaedel et al. 2010; Shen et al. 2008; Shen et al. 2009; Zheng et al. 2007).

Survey research, which relies on the ability of the survey taker to assess the OC contributions, has been problematic in several ways. First, the survey taker has opinions and feelings that might bias the survey results. Second, the survey instrument is inexact. Typically a survey relies on measures such as 5 point or 7 point Likert scales. This inexact measure introduces measurement error on the part of the survey taker (i.e. whether to put 5 or 6 for a question) and by the researcher interpreting the data. Third, the survey respondents may be a self-selected group that biases the data. Those people who answer surveys may be representative of a group of people that correlate with people likely to respond to a survey in the dataset. On the other hand the non-responders of surveys may be a group that is not represented in the dataset. For example lurkers tend not to post or reply to posts. Lurkers would likely not answer a request to fill out a survey. Thus lurkers may be a group that is not represented in a survey study. I propose using a bibliometric measure that looks at influence by quantitatively looking at actual replies and views (i.e. reads or displays to the computer screen).

The Hirsch-index was proposed by Hirsch in 2005 and is defined as "A scientist has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each." In order to find the h -index of an author you would gather all the publications by said author and the citations to those publications. Then you would list the publications in descending

citations order. When the i -th publication from the top has less than i citations you have found the h -index for that author.

The Hirsch index has been proposed as a way to look at OC postings in the past (Gomez et al. 2008; Laniado et al. 2011). In Gomez et al. (2008), they look at the discussion threads on Slashdot, a popular website with technology news. The researchers take the postings and developed a “measure to evaluate the degree of controversy provoked by a post” (Gomez et al. 2008). They do this by taking each initial post and create a tree from the initial post. The branches of the tree are responses to the initial post and the subsequent responses to the responses. The Gomez et al. h -index measures the maximum nesting level h of the tree where h “has at least h comments, or in other words, $h+1$ is the first nesting level h which has less than h comments” (Gomez et al. 2008).

The Laniado et al. (2011) paper looks at wikipedia discussion pages. They try to characterize the discussions on the comments section of wikipedia. Laniado et al. (2011) do this by looking at the depth of the discussion on an initial comment as well as the breadth of the discussion using a h -index like measure. Laniado et al. (2011) construct trees out of the discussion, similar to Gomez et al. (2008). Laniado et al. (2011) create the tree using the response comments, which become child nodes for the initial discussion entry. For Laniado et al. (2011) the “ h -index of the tree is then the maximum level, for which the corresponding number of nodes is greater or equal to the level number” (Laniado et al. 2011). This is similar to the Gomez et al. (2008) version of the h -index. One key difference between these authors definition of the h -index from the original h -index is that the unit of measure for these discussion pages is the initial posting entry and measures the h -index at the level of posts and not the authors of these entries. They are able to find controversial and impactful initial posts but they cannot identify the individual contributors that are responsible for these posts using their h -index.

I use the h -index measure to define the influence for a contributor to the OC. I interpret citations in the OC to be replies to posts made in the OC. My use of the h -index is in line with the initial definition of the h -index in that I am trying to find the authors of impactful posts. I am not interested in finding the impactful postings but rather want to find the authors that are creating interesting content that is being read and commented on by a large number of other users of the OC. I am different from the use of the h -index of Gomez et al. (2008) and Laniado et al. (2011) in that those authors have used it to measure influence in individual posts, while I kept the original Hirsch (2005) h -index unit of measure of the influential author. The difference that I have between the Hirsch (2005) definition is that I use comments instead of citations to measure the impact of the author of the posting in an OC.

I also have the distinct advantage over bibliometric studies in that I can find out who has viewed (or at least displayed onto their computer) posts from other users in the OC. So in addition to the replies (citation) I have a second h-index measure that looks at views. I define these as being the hr-index (replies) and hv-index (views). So the hr-index is defined as a contributor has index hr if hr of his/her N_p thread starts have at least hr comments each, and the other ($N_p - hr$) thread starts have no more than hr comments each. I also define the hv-index similarly as a contributor has index hv if hv of his/her N_p thread starts have at least hv comments each, and the other ($N_p - hv$) thread starts have no more than hv comments each.

2.2.4. Social Network Analysis

I also utilized several measures from social network analysis (SNA). I used the degree, closeness, and betweenness centrality measures. Centrality in SNA refers to the notion that in a network there are nodes that are central to the network. That is in a social network there are individuals that are central to the group. If a person is central to the group that means there are more people connected to that person in the group. Degree centrality is this simple count of how many connections you have in the group (Freeman 1977). The theoretical fully connected person will have $n-1$ connections (where n is the number of people in the group) or connections to everyone in the group except for herself. When someone is fully connected the distance between them and everyone else in the group is one (Acedo et al. 2006; Eaton et al. 1999; Henry et al. 2007; Otto et al. 2002; Polites et al. 2008; Vidgen et al. 2007). Degree centrality is a measure of how many people you know in the network.

Closeness centrality is a measure of how distant you are from other individuals in the group. The distance is measured by the minimum number of connections to reach that other person. So for one individual, you would first find the distance between them and all $n-1$ other individuals in the group. The closeness centrality then finds the mean of these distances (Polites et al. 2008; Vidgen et al. 2007). The closeness centrality is a measure of how close you are to the rest of the network.

Betweenness centrality is a measure of how much you are an intermediary between two people on a network. If you are part of a minimum distance path between two people you are the "friend that knows a friend". So people need you to connect to whomever they want to connect to. You are one of the crucial intermediaries to allow connections between two people. The betweenness finds all the shortest distances that go through a person and then sums these distances and divides by $(n-2)(n-1)$ (Vidgen et al. 2007). The betweenness centrality is a measure of how important of an intermediary you are to connect individuals in the network.

2.3. Research Context

Parts of this dissertation focuses on the case of BigUFans, an independently owned and operated website that covers and provides an online community of like-minded fans of a large university in the United States. The site owner has been looking for better measures of user value besides the simple count of posts. The owner realizes that maintaining a core group of users who participate regularly is the key to success in an online community. Other members typically wait until these core members post before they themselves reply. However, the owner also realized that it was not just the quantity of posts that identified the core user group. Instead, in OC's like BigUFans, the quality of the post and the reputation of the member are factors in how engaged other members become on a particular topic. BigUFans is representative of many similar sites, which court advertisers to a hard-core, niche group of dedicated followers. It is through advertising that the site has, over nearly 15 years, earned enough revenue to keep the site online and improve its server capabilities, which, in turn, have made the site more usable. However, it is the vibrancy of the community that makes it relevant. A better understanding of which users are key to this vibrancy is the goal of this research.

BigUFans.com is an OC website for the school Big University. Big U. is a large tier I research institution in the southeastern US. Big U. is a public state university that has been around for 150 years with over 30,000 students. The OC is in particular geared towards the athletic teams of Big U. where football is the major sport. Big U competes in the NCAA Division I in many athletic disciplines including the major sports of football, basketball, and baseball.

Currently BigUFans.com has four major competitors. BigUFans.com was created in 1999 and is the largest free OC out of five competing OC's. The OC makes revenue by advertising sales. Thus gaining activity on BigUFans.com is an important revenue generating activity. The four other competitor sites have some aspect of paid areas in their OC. Anyone may anonymously browse the contents of BigUFans.com but in order to post or reply to forums a user must register.

There are six administrators and 17 moderators of BigUFans.com that work to keep the topics on the forum of BigUFans.com relevant and free of spam and incendiary comments. Registered members can become moderators as they become more active in the community but only at the invitations of the owner of the site. Because of the work of the staff, one of the major advantages of BigUFans.com over their competitors is that the community feels the OC has more positive messages that are free of advertising, spam, flame wars, and off-topic postings.

The majority of activity on BigUFans.com comes from discussion about the football program. During the 2011-2012 season monthly activity saw an average of 2.9 million page views, 1075 new posts, 167,000 unique visitors, and 89 topics started. Daily activity included 1,017 signed in users and 5 new users. A typical visit consisted of a viewing of 5.2 pages. The demographics were typically male (92%), college graduate (72%), and of middle to upper income (71%). During the offseason prior to the 2011-2012 season, there were 2 million page views a month. Currently BigUFans.com gives different level user status automatically to users that post a certain amount, regardless of how many users read their messages. Currently BigUFans gives the following user status (along with minimum posts required); Scout Team (0 minimum posts required), Third team (13), Second Team (50), First Team (250), All-Conference (1000), All-American (2000), and Hall-of-Fame (5000). Currently there are 20470 Scout Team members, 1363 Third Team members, 1086 Second Team members, 590 First Team members, 179 All-Conference members, 121 All-American members, and 47 Hall-of-Fame members. The Scout Team is basically given to anyone that signs up for the service. In addition, there are six Administrators and 17 Moderators. These are given to users that have established themselves as responsible members (i.e. no spamming of message boards) and are asked to join these higher ranks.

2.4. Papers

The following papers were presented at conferences, submitted to journals or are in the process of being submitted to outlets. Currently the research has produced five conference presentations.

2.4.1. Investigating the Impact of Offline Events on Group Development in an Online Sports Community

This paper was presented at the Southern Association for Information Systems conference in 2011 in Atlanta. This paper looked at the literature on online communities and found those communities that grow successfully have members who create and share common bonds and common identities. The goal in this research-in-progress paper is to better understand the impact that shared experiences have on user activities in online communities. The research used PLS in a preliminary analysis of the relationships between page views at the online community BigUfans.com and real-life events that affect the online community's members. The preliminary results suggest that our model accounts for 60.4% of the variance in page views on the site and 27.2% of the contributions (e.g. message board posts) to the site.

2.4.2. Online Community User Influence: A Study Using User Status

This paper was presented at the Southern Association for Information Systems conference in 2012 in Atlanta. This research in progress will try to understand the relationship between user influence and status of users in an online community. For online communities a major issue is how to continue to grow and sustain the viability of the community. Oftentimes online communities will gain popularity and then lose that popularity due to stagnation in the number of new users and loss of interest by current users. This research tries to understand how the status of the user plays in the influence of the high level users and the vibrancy of the online community. We measure the influence of the user by using a pseudo Hirsh index to measure the interest of the users posts and use the status of the user in the online community to gauge the user status in the online community.

2.4.3. What Drives User Contribution in an Online Community? A Study in Contributor Influence and User Status

This paper was presented at the America's Conference on Information Systems conference in 2012 in Seattle. Online communities (OC's) depend on shared interests and user interactions mediated by technology. Successful OC's find ways to encourage these interactions to grow communities. Many OC's have influential users that help grow the community by their very presence and contributions. However, the process for identifying users having the greatest impact is not trivial. This study offers a new method for identifying these influential users through the creation of modified Hirsch indices, which improves upon the current method of using contribution counts or a survey method of polling other users. We validate the new measures against user status and then analyze the measures by correlating them against postings, thread starts, and views and replies to the thread starts for a shared interest OC.

2.4.4. Online Community of Influence: A Study Using the Hirsch Metric and Social Network Analysis

This paper was presented at the America's Conference on Information Systems conference in 2013 in Chicago. This study looks at small to medium sized online community (OC) and tries to identify ways to measure impact of the contributions of the users of the OC. OC's are dependent on contributions of their users to maintain the health of the OC. Measuring the health of an OC by identifying those users that have most influence and thus create more activity and finally more people the visit the OC is an important activity to the stakeholders in the OC. In order to measure these high end users we

are extending previous research to include a two-part measure. First, using the Hirsch metrics to measure the productivity and impact of user contributions and second, using social network analysis to see those users that have high centrality measures in the network of posters and readers of the OC. This study looks at one University sports fan site to measure the influence of their users and found some correlation between the Hirsch measures and the centrality measures.

2.4.5. Measuring New Member Contribution in an Online Community: A Preferential Attachment Approach

This paper has been submitted for review at the International Conference on Information Systems (ICIS) in 2015 in Fort Worth. The online community (OC) is a popular form of specialized knowledge transfer, where geographically dispersed users can form a community to share ideas, send messages, and forge online friendships. OC's tend to have a life cycle, with the birth and growth of the OC followed by a stagnant stage where users stop posting and the OC eventually dies due to inactivity. One way that an OC can stay vibrant is to encourage new contributors. In this research in progress, we evaluate new users of an OC for preferential attachment, where new users tend to interact and post on thread starts of authors that are well known and respected in the online community. We look at the existing users' posting levels to measure veteran users' impact and measure new user behavior compared to veteran user of the OC. We operationalize our method using a college sports team fan site.

2.4.6. Measuring Member Contribution Impact in an Online Community: A New Hirsch Index Inspired Approach.

This paper is the first journal level submission from this research stream. The paper was submitted for a special issue in ISR in 2015. This is a continuation of the work presented at the SAIS and AMCIS conference in 2012 (see section 4.2, 4.3). The online community is a popular form of specialized knowledge transfer, where geographically dispersed users can form a community by sharing ideas, sending and posting messages, debating topics, and forging online friendships. One of the problems with online communities is that they tend to have a life cycle, where there is the birth and growth of the online community but then there is a stagnant stage where users stop posting to the online community and the community eventually dies due to inactivity. Trying to extend the vibrant growth stage of an online community is a relevant topic for any administrator of an online community. One way that an online community can stay vibrant is to encourage contributions.

In this research, I propose using a non-invasive bibliometric measure using the Hirsch-index methodology as a way to identify high-level contributors. I proposed that those users who are high in

perceived identity verification in the form of replies and views to threads started would have high contribution in the form of the Hirsch index measure. I ran the data on a college sports fan site. I also analyzed the data to identify high-level contributors to the online community. Finally, I demonstrated how this new measure can be used using a real online community.

2.5. Discussion and Implications

There are some benefits and limitations to this research.

2.5.1. Benefits

The use of user status would be given to a user once they post a certain number of posts since the time they started posting in the OC. While this is great to encourage users to come back and post often, I find that these statuses are artificially inflated. Since the measure is a product of postings which is taken over all time the OC has been in existence I feel that a more democratic view of looking at a measure that takes the behavior of other users of the OC would be a much better measure of influence. A dataset such as SNA measures sees the behavior of the readers of the OC. A dataset such as the h-indices measures the behavior of other users in their viewing and replying to other users.

In addition, I find that the aggregate measure is a much better measure in that it can easily look at a short timestamp of data to find who is relevant in the past short time period, whether it be a few weeks or months. I also feel that using the aggregate measure I can find different categories of users. I have started to identify some of these such as 'controversial messenger', 'lurker', or 'information provider'. I feel that future research can use these types of categories to allow the owner of the OC to target different users for possible behavior changes. Maybe try to get the lurker to contribute more. Maybe see if the 'controversial messenger' is creating incendiary posts and target them for possible disciplinary action, such as being banned from the OC. Or maybe encourage the 'information provider' to put out more posts that can stir up more discussion.

2.5.2. Implication for Practitioners

I feel this measure will be a better way to evaluate users for owners of OC's. The current method that most OC's use with rating people based solely on the amount of postings or content creation does not give a measure of impact. For practitioners, the interest should be on how impactful certain

users posts are. This means that a measure like mine should be used for encouraging users that participate.

Another implication is that different types of users start to emerge using the measure. Users such as 'controversial messenger', 'lurker', or 'information provider' start to emerge. This allow the practitioner to identify these different types of users and target encouragement depending on the type of users and how the practitioners want the users to improve the vibrancy of the OC.

This also allows a total analysis of the OC with the possibility of using archival data on the OC. Compared to using survey methods, which can bias the output due to response trends in surveys, using the non-invasive measures allows all data to be collected.

2.5.3. Limitations

While the measures identified were an improvement over past methods, there are some limitations. First the dataset used is not generalizable to other OC's. The current data was taken on a free, university sports fan OC. The generalizability not only to other university fan OC's or other sports fan OC's but to other types of OC's comes into question. A sports fan OC also has the advantage of having topics of conversation that gets updated every season and game-to-game. For another type of OC such as a computer gaming OC, once the community members finish playing the game and conversations about these games have been saturated, and there are minimal topics of conversation to continue the OC. For a sports fan OC this is different as many possible topics of conversation can be planted from game to game. New players might be recruited, new injuries might occur, and game strategies can be discussed from game to game. So the generalizability of the dataset can be questioned in several ways. In addition I used a six-month period and were able to get a very large dataset. The type of OC to generalize to might be different from an OC that might not have this type of activity week to week.

Another limitation may be the use of replies and views. As the hr and hv indices have been derived from the h-index, the replies and views for hr and hv, respectively, were substituted for the citations counted in the h-index. One can argue that replies to an OC thread is nothing compared to the work that goes into a citation. Worse yet, the views are even easier for an OC member than the replies. One can in addition argue that the hv might be less important, since the views do not necessarily mean that one has 'viewed' the message, it only means that the message was clicked on and displayed on the computer. Another issue that arises from views being easier than replies is that since the views are easier the hv-index is considerably higher than the hr-index. So while the hv-

index might mean less due to the ease of views over replies, the hv-index may have larger numbers, which makes comparing the two more difficult.

A third set of limitations comes from the use of the h-index methods in the measures. The h-index has been criticized for several items. The main problems cited with the h-index are the fact that the h-index is harder to gain for more junior researchers, the h-index does not measure recent research, and the h-index does not take into account of the large citation hits. The first two criticisms of the h-index are intertwined. One fact about the h-index is that it can never go down. This is due to the fact that once you gain a citation, that citation never disappears. So if a researcher gets an h-index of 1, that number will never go back down to zero. This ties to the fact that a junior researcher has a harder time garnering an h-index than a more seasoned researcher, since recent publications have a difficult time to gain an h-index. The h-index is tied with citations and for recent publications; it takes time to garner those citations. One cannot predict the citation trend down the road so it becomes harder for junior researcher to gain an h-index.

The third issue with the h-index is that the papers that garner a large number citations, the “one hit wonders” tend to be overlooked by the h-index. The h-index is a measure of both publication productivity and the impact of those publications via citations. If there is a researcher that published two papers over their career but those two papers received thousands of citations, the h-index for that researcher will not be able to get higher than two. Due to the definition of the h-index, which counts the number of publications, the maximum h-index that one can get is limited by the number of publications that the researcher has published. In the same vein for the hr and hv index, are limited to the number of thread starts that an OC member has begun.

2.6. Table of Research

The following table shows the publications that have been presented and are currently in the pipeline. A total of five conference presentations have been presented, one conference submission is in review and one submission to a journal has been made. Future research will continue on this stream of research. The table below summarized the research stream.

Table 2-1. Table of Research

Title	Objective	Contribution	Theory	Methodology	Data Collection Period	Results	My Contribution	Future	Venue
Investigating the Impact of Offline Events on Group Development in an Online Sports Community (Ch. 3, section 4.1)	Find the relationship between real life activities and the activities on an online community	Online community activity tied to real life activities	Common Bond and Common Identity	Conceptual model using bootstrapping and PLS	Three months (Aug. 1, 2009 to Jan. 31, 2010)	Real life activities have a significant effect on contributions to an OC	Data analysis, majority of writing, presentation at conference	Data collection and journal submission	SAIS 2011
Online Community User Influence : A Study Using User Status (Ch. 4, section 4.2)	Does user status in an OC really correspond to user level contributors	Validity of hr and hv indices to measure contribution to OC	Bibliometrics	Use hr and hv to measure influential postings and see if user status corresponds	No data taken, research proposal only	Research proposal only	Majority of writing, presentation at conference	Precursor to the AMCIS 2012 paper (Ch. 5) and journal submission (Ch. 8)	SAIS 2012
What Drives User Contribution in an Online Community ? A Study in Contributor Influence and User Status (Ch. 5, section 4.3)	Find out if hr and hv are better predictors of user level contributors	Empirical validation of hr and hv indices	Bibliometrics	Correlation analysis of hr, hv, postings, thread starts, views, and replies over a one and two year period.	Three years, the entirety of 2009, 2010, 2011	User status, hr, and hv predict 1 year of OC views and replies but hr and hv are better than user status at predicting over a two year period	Primary author, majority of writing, data analysis, writing of Java program, presentation at conference	Journal submission (Ch. 8)	AMCIS 2012

Online Community Influence : A Study Using the Hirsch Metric and Social Network Analysis (Ch. 6, section 4.4)	See if social network centrality and h-index correlate for user contribution in OC	Addition of SNA to research stream	Bibliometrics and Social Network Analysis	Correlation analysis of SNA measures and h-indices	Six months, (Aug. 2012 to Feb. 2013)	Support shown for hr, hv to degree and betweenness centrality but not closeness centrality	Majority of writing, data analysis, writing of Java program, 45present45at ion at conference	Data collection for journal submission	AMCIS 2013
Measuring New Member Contribution in an Online Community : A Preferential Attachment Approach (Ch. 7, section 4.5)	Try to explain health and vibrancy of OC by focusing in on new users behavior and trying to get those user to become 45ven level contributors	Use of Preferential Attachment to explain new user behavior in OC	Preferential Attachment, network analysis	Analyzed behavior of new users to explain propositions	One year for new users (2014) 45for veterans 45 data collected for 15 years (1999 to 2014)	Support for new users to start new threads, 45vents45 users to reply more, and to reply to 45vents45 users more	Primary author, majority of writing, data analysis, writing of Java program	Awaiting 45vents45ati acceptance decision	ICIS 2015 (in review)
Measuring Member Contribution Impact in an Online Community : A New Hirsch Index Inspired Approach (Ch. 8, section 4.6)	Try to find a better way to measure member contribution in an OC	Explain contributions of users with hr, hv, and SNA and an aggregate 'Rank Sum' measure	Bibliometrics, SNA	Calculated the measures and analyzed the results	Six months (Oct. 23 2012 to April 21, 2013)	Rank Sum seems to measure 45ven contributors. Betweenness 45vents45ati with hr, hv, and closeness has no bearing on measures.	Primary author, majority of writing, data analysis, writing of Java program, 45present45at ion at conference	Journal submission	In process

2.7. Conclusion

I commenced this research to find new methods for OC owners to find users that contribute to the vibrancy of the OC. The problem that OC's face is the lifecycle of the OC and the possibility of losing relevancy and timely information and becoming a statistic in the graveyard of past OC's. I propose an aggregate measure using SNA and bibliometric measures that are dependent on the behavior of the OC members. These measures are shown to be more descriptive and insightful than the current status quo of using user status or posting frequencies.

I ran data on BigU an online community for a university sports team. The data was taken over a six-month period and showed the diverse nature of the top contributors to the OC. I was able to identify several categories of users that can be helpful for an owner of the OC to identify high-level users.

Chapitre 3

Investigating the Impact of Offline Events on Group Development in an Online Sports Community

Abstract

The literature on online communities suggests those communities that grow successfully have members who create and share common bonds and common identities. Our goal in this research-in-progress paper is to better understand the impact that shared experiences have on user activities in online communities. We use PLS in a preliminary analysis of the relationships between page views at the online community BigUfans.com and real-life events that affect the online community's members. The preliminary results suggest that our model accounts for 60.4% of the variance in page views on the site and 27.2% of the contributions (e.g. message board posts) to the site.

Keywords

Online communities, Website design, Group development theory, and Group Collaborations

3.1. Introduction

Online sports communities continue to grow in popularity and sophistication as a means for geographically dispersed followers to stay connected. Online communities (OC's) are often used as a source of knowledge and a repository of information from which members draw on their real-life experiences. These OC's offer a way for people with common interests to actively participate in sharing of information, debating relevant (or irrelevant) topics, forging online connections, and building a sense of being part of a vibrant virtual community. But for these communities to exist there needs to be active participation by the members.

College sports message boards are particularly representative of popular OC's. For example, on February 7, 2007, more than 70 million visitors logged onto popular college fan sites owned by Rivals.com to discuss college football (Skretta 2007). This level of interest is also present, albeit on a smaller scale, in fan communities such as BigUfans.com (a pseudonym), the subject of the present study. Owners of fan websites that operate for profit are typically interested in increasing the number of page views in order to increase their revenues from the site. Toward that end, this study begins a line of action research to investigate how site owners can increase page views and thereby revenue for their site. In this paper we investigate the effect that events related to the community interest have on page views of the site. The object of such research would be to provide insights to the owners of those sites to assist them in maximizing page views and revenues.

The rest of the paper is structured as follows: First, we briefly describe group development theory. Second, we present the research questions. Third, we present the empirical context for the research study. Fourth, we present the research method and dataset. Finally, we give a preliminary analysis of the data and offer concluding remarks.

3.2. Literature Review

Many studies are conducted on online communities (e.g. Clark et al. 2007; Harper et al. 2007; Iriberry and Leroy 2009; Ren et al. 2007; Sassenberg and Postmes 2002). One major concern for OC's is that they may start out strong but eventually a once-relevant niche loses momentum and the community loses its vibrancy. This loss of vibrancy results in a loss of page views and therefore revenue. Identifying why this occurs and taking counter-measures can help the site owner maintain and increase page views and thereby revenue. Studies have examined reasons for decreases in community member activity (Harper et al. 2007), OC life cycles (Iriberry and Leroy 2009), moderation techniques impacting community growth (Ren and Kraut 2007), and indications that virtual

communities may not be entirely virtual (Rothaermel and Sugiyama 2001). Group development theory (Back 1951) has the concepts of common bond groups and common identity groups. Back identified three reasons someone would want to be in a group: (1) personal attraction between members (common bond); (2) personal attraction to the group activity (common identity); and (3) prestige of being in the group. The notion of common bond and common identity translates to the online world as well (Postmes and Spears 2000; Ren et al. 2007).

Common identity. Common identity is where “members feel a commitment to the online community’s purpose or topic” (Ren et al. 2007, pp. 381). Reasons for common identity are social categorization and interdependence. With social categorization members of a group exist due to a shared identity such as a group dedicated to those suffering from cancer. Interdependence occurs when the group exists for a common purpose or some goal (Ren et al. 2007).

Common bond. Common bond “implies that members feel socially or emotionally attached to particular members of the online community” (Ren et al. 2007, pp. 381). Reasons for common bond are social interaction, personal information, and similarity. Social interaction occurs during messaging or group forums. Users may find attraction to a person in the community through these interactions. The attraction might be due to favorable ideas shown during posting or support shown by replying positively to a post. Similarity is when one finds another person that might have similar tastes, ideas, notions, or features. Similarity can be a reason why a user continues to be in an OC (Ren et al. 2007).

3.3. Research Questions

The importance of this research is to explore how an interest-based website can increase page views through creation or taking advantage of major events. Management of events in the life of the community can increase activity on the site which impacts revenue to the site. By identifying how the website members behave, this study can identify when and how a website owner might address different events. Koh et al. (2007) suggests that these communities must consider ways to sustain interest, viewership, and contributions in the community. Community size also is an important factor in the community’s survival. Therefore, it is important to understand how to use events that occur offline to create growth within the OC. Thus, we have several research questions we wish to explore:

RQ1. How do real-life vs. website events in interest-based website visitors’ lives influence user participation?

RQ2. How can interest-based websites take advantage of the real-life common events to increase user participation?

We believe that the reason members are active in sites like BigUfans.com is both common bond and common identity based. With common bond, the attractiveness of the community is most likely social interaction and similarity. Since personal information disclosure is not required, members participating due to personal information are likely low. While not required, some users do disclose additional personal information such as cell phone numbers when someone has an extra ticket to a sporting event that they want to give to another fan of Big U. Optional information disclosed in member profiles includes geographic location, occupation, and personal interests.

With common identity the site can be categorized as being both social categorization and interdependence. Many members of the site are alumni of the university, live in the proximity of the university, or have been lifelong fans, thus giving a common identity via attendance to the university or living in the community. Most members of the site are fans of the university and thus have a common goal (interdependence) cheering for their university at sporting events and thus wishing team victory.

While some groups may have prestige, the website currently under study does not have any prestige associated with its membership. Due to this fact we did not use Back's (1951) third reason for someone wanting to be in the group.

3.4. The Empirical Context

BigUfans.com is an independent website and home to the largest free gathering place for fans of Big U's athletic teams. Big U is a public university in the southern United States with enrollment of over 25,000 total students and over 100,000 living alumni. Big U competes in the NCAA at the highest levels in many sports. BigUfans.com is a free interactive social networking community created by an alumnus (and one of this paper's authors) of the university. The website is not affiliated with, supported, or endorsed by Big U. This OC is typically most active in months related to the college football season (September through January). During a 180-day period of 08/01/2009 to 01/31/2010, the site had an active membership 3,079 users that accessed the site during this period. As of 4/15/2010, there was a total (active and inactive) membership count of 21,392. Approximately 20% of all visitors on any given day logged-in as registered members. The other 80% were non-members and members who did not login and thus were unable to contribute. 'Anonymous' postings by a visitor not logged in were not allowed. Users were not required to disclose any private information to the public or members of the community.

During the 2009 college football season, the site averaged 2,094,226 page views and 151,286 unique visitors per month. Visitors spent on average 8 minutes 41 seconds on the site per visit during the season and viewed an average of 6.6 pages per visit. Each day an average of 1,075 registered members logged into the site. Each day during the college football season, on average, 13 new members registered, 73 new topic threads were created, and 1,022 new posts were created. BigUfans.com is unique (as compared with the vast majority of large sports fans websites) in that all content and message boards are freely available to all visitors. Approximately 90% of the revenues for BigUfans.com come from click-throughs to advertisers. Merchandise sales account for the remainder. Click-throughs is a function of having many people view the site. Users regularly visit the site to read and comment on news and interact with other fans. We believe that the common experiences of a Big U fan tend to generate page views surrounding major events in the life of a Big U fan. There is a base level of approximately 40,000 page views regardless of events (e.g. 30,000-40,000 that reflects a general interest in BigUfans.com. Events cause the number of contributions to increase, which, in turn generates additional contributions and page views.

3.5. Research Model and Variables of Interest

Page views are one measure of success for websites. For interest-based communities like BigUfans.com increasing the page views is an important goal and is linked to site revenue. Other measures that OC designers closely monitor are sign-ins and sign-ups. Common events in the life of the community and its members can influence the page views for the site. Thus, our conceptual model is presented in Figure 3-1.

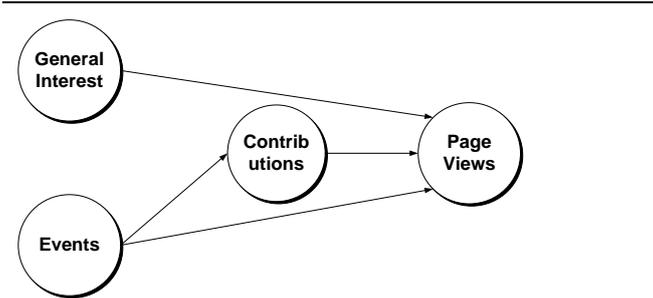


Figure 3-1. The Conceptual Model

Dependent Variable. Page Views are measured as the number of views of a page recorded each day on BigUfans.com.

Independent Variables. Four types of events variables were developed by creating a dummy variable (1/0) for each day. Specifically, event data is indicated two days before an expected event occurred, the day of the event, and the two days following the event. Where an event occurred that was unexpected, the event and the two following days were coded as being for that event. For example, if an expected major event occurred on a Saturday, Thursday and Friday would be the prior-to-event, Saturday would be the event, and Sunday and Monday would be post-event.

Major Events consist of football games against the biggest rivals, post-regular season games, and football recruit signing days. In our research, this includes games against the three biggest rivals for Big U. Games played after the season, such as conference championships and bowl games, and head coaching changes or NCAA penalties are considered as major events. The variable is dichotomous for every day in the data set with 1 being a major event occurred on the day and 0 indicating no major event occurred on the day. *Minor Events* consist of games against smaller rivals and other, less impactful school events. These consist of all other games that Big U plays, announcements by the school regarding player eligibility, or relatively minor announcements about Big U regarding players or recruits. The variable is dichotomous for every day in the data set with 1 being a minor event occurred on the day and 0 indicating no minor event occurred on the day. *Adverse Events* include losses to major rivals or unexpected losses to minor rivals. Losses to major rivals and unexpected losses to minor rivals are of significant interest to fans. The variable is dichotomous for every day in the data set with 1 being an adverse event occurred on the day and 0 indicating no adverse event occurred on the day. *Expected Events* include wins over minor rivals. The interest of the fan base in Big U is atypically high during football season, but less so for expected events. The variable is dichotomous for every day in the data set with 1 being an expected event occurred on the day and 0 indicating no expected event occurred on the day. *Articles* indicates publication of an article by a popular BigUfans.com columnist. This variable was coded 1 on the publication day of the article and 0 for all other days.

Mediating Variable. Contributions are the number of member posts contributed on the website each day. These are measured using website data from the BigUfans.com's message board control panel. Contributions include all new posts created by members each day across the website and are not limited to only posts in the football board.

Control Variable. Days indicates number of days since the start of collection of the website data. This was a control variable created to parse out the effect of expected growth in the site over time.

3.6. Hypotheses

Membership provides the privilege of posting on the site. Visitors are not required to be signed in to read content and postings on the site. However, since the sites primary purpose is to create a place for fans to interact, the dependent variable should increase as the participation in the site increases. Thus, *H1: The number of contributions mediates most of the influence of events on page views.*

In the life of the site, major events will provoke members to participate in the life of the site. In fact, the level of activity of the site, as measured by page views, sign ins, sign ups, new threads, new posts, and unique visitors, is expected to increase during the period surrounding major events. Thus, *H2a/b: Major events generate a greater number of page views (contributions) than other events.*

While minor events are not as special as major events, the minor events are still going to generate an increase in the activity of the website compared to non-event days. Because of the common identity that members and non-members share, we expect to see greater activity on the site when adverse events occur. Thus, *H3a/b: Minor events generate a greater number of base-level page views (contributions) than non-event days* and *H4a/b: Adverse events generate large increases in page views (contributions) above those in major or minor events.*

We also expect to see a difference between the common bond and common identity events as well. For both events we expect to see increased activity but not the same dependent variables will be affected. For common identity events we expect to see an increase in contribution across all dependent variable measures. Thus, *H5a/b: Common identity events generate more page views (contributions) than non-event days.*

For common bond events, one must already be familiar about the person (in our case the staff writer) in order to feel some attraction. So new user events will not play a factor for common bond events. Thus, *H6a: Common bond events generate more page views (contributions) than non-event days.*

3.7. Preliminary Analysis

Method. To test the hypotheses, we examined the relationship between page views at BigUfans.com and the occurrence of certain events. We collected usage data from BigUfans.com and time periods measured by days and proximity to associated relevant events categorized as being positive/negative and minor/major events.. We used Partial Least Squares (PLS) methodology as implemented in SimplePLS (Ringle et al. 2005). PLS is a second generation data analysis technique (Gefen et al. 2000) which tests not only the structural model but also the measurement model in a single analysis rather

than two unrelated analyses as in the first generation techniques. Additionally, PLS is able to identify path loadings across the entire model in a single run as opposed to multiple runs required using regression techniques. This results in a more rigorous analysis than using factor analysis and regression alone (Gefen et al. 2000, p. 24). Similar to regression analysis, PLS seeks to show rejection of a null hypothesis of independent variables having no effect on the dependent variable while accounting for a significant amount of the variance in the dependent variable (Gefen et al. 2000, p.27).

Data. The data set was made available by the owner of the site (one of the researchers) to the research team. Data includes directly measured data from analytical tools located on the site, including Google Analytics, 24/7 RealMedia, and vBulletin message board software. Google Analytics and 24/7 RealMedia place a small code on each page delivered to viewers of the site and then collects various types of data. Data collection occurred during the timeframe of July 2004 through May 2010

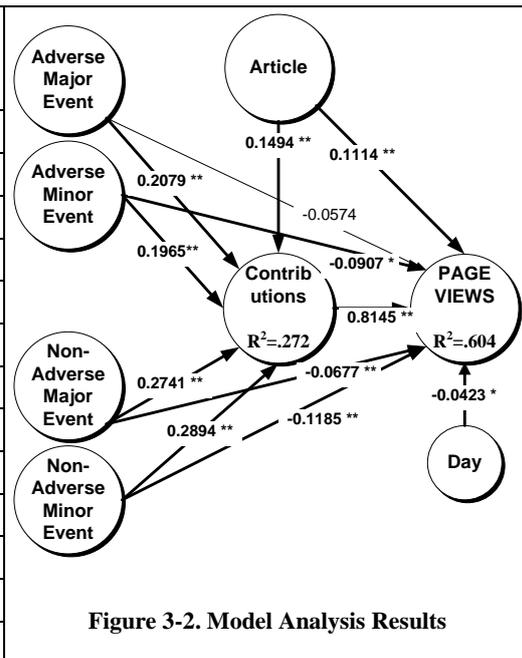
Analysis. The data collected was analyzed using the SmartPLS Software (Ringle et al. 2005). A structural model was created and the analysis was performed by running a PLS analysis with using the path weighting scheme and other standard settings. To get the path significance, a bootstrap analysis was performed by creating 1000 item samples and then collecting 200 samples. Since the data is single indicator variables collected from an automated system, no testing of the measurement model is required. A structural model was developed in PLS and the model accounted for 60.4% of the variance in page views and 27.2% of the contributions. Both are very high values. The results of the PLS analysis is presented in table 3-1 and figure 3-2.

Path Analysis. The path analysis shows that all of the paths tested were significant except for the path from adverse major events to page views (see Table 3-1). The significant variables had a direct effect on page views and on contributions. However, the model shows negative significant direct paths for the events to page views. This is indicative of suppression. In this event it indicates multiple contradictory processes are occurring. For, example it is possible two different groups have different responses to events. One group perhaps may not view many pages, but when an event occurs it causes them to increase their page views, while another group has high page views, but when an event occurs, it causes them to stop viewing the site temporarily. The suppression effect indicates that additional research and theorization is necessary to specify all of the entities involved in this model (Shrout and Bolger 2002). To evaluate H1, H3, and H4 additional research is required.

The entities involved in this model (Shrout and Bolger 2002). To evaluate H1, H3, and H4 additional research is required.

Table 3-1. Results of Bootstrapping Analysis

Path	Original Sample (O)	Sample Mean (M)	t-value
Note: * = p<0.05, ** = p<0.01			
Adverse Major Event _ Page Views	-0.0627	-0.0574	1.6266
Adverse Major Event _ Contributions	0.2155	0.2079	4.2470**
Adverse Minor Event _ Page Views	-0.0919	-0.0907	2.3081*
Adverse Minor Event _ Contributions	0.1976	0.1965	3.3500**
Contributions _ Page Views	0.8115	0.8145	28.0585**
Article _ Page Views	0.1131	0.1114	4.5121**
Article _ Contributions	0.1466	0.1494	4.2490**
Non-Adverse Major Event _ Page Views	-0.0660	-0.0677	2.0270*
Non-Adverse Major Event _ Contributions	0.2736	0.2741	5.5468**
Non-Adverse Minor Event _ Page Views	-0.1197	-0.1185	4.6478**
Non-Adverse Minor Event _ Contributions	0.2874	0.2894	8.4160**
Control Variable: Ctrl_Day _ Page Views	-0.0432	-0.0423	2.1399*



The betas of major events and minor events are approximately the same for both their effects on contributions and on page views indicating that their effects on both of those are about the same. Thus, H2 is not supported. For the publication of an article, the results show a partial mediation effect. The total standardized effect of the article on page views is .2321. The beta of .1131 for the article variable yields a partial mediation effect of approximately 49% (Shrout and Bolger 2002). This indicates that the effect of the articles is split; half drives page views to the article itself and the other half drives contributions and comments about the article. The control variable, Day has a significant, negative beta indicating that over time, general interest in the site is declining which contradicts the observed growth in page views over time. This could be a suppression effect with other variables or it could mean that general interest in the site is declining over time. However, a core group is strongly committed to the site and actively viewing and posting contributions on the site. The high beta for contributions to page views shows the major contributor to page views is the number of postings on the site. The number of postings is driven by articles posted as well as by events that occur to the team over time.

3.8. Conclusion

The goal of this research is to contribute to our understanding of how interest-based websites can take advantage of events that affect the website’s visitors. Preliminary results show that events that occur in the life of the community have a significant effect on contributions to the site and, through contributions, the page view count. The effect of articles posted on the website is partially mediated by the contributions in their effect on page views. Our work builds upon prior work on group

development theory in two significant ways. First, this research applies to interest-based OC's. A general commercial site with no natural interested-based community will have different dynamics than an interest-based site. Second, our study uses group development theory for understanding how these shared experiences impact the life of an online sports fan community. The increase in interest-based OC's has provided users with common identities and common bonds the ability to share their experiences together, even while geographically dispersed. Many OC owners are interested in creating sites that grow and generate income. Understanding target areas to take advantage of offline events can be an important contribution.

We ran PLS analysis on data collected from a university sports fan site. We found support for the hypotheses (H5 and H6) that Common identity and Common bond events generate more page views than non-event days. We found weaker support for the hypotheses (H1, H3, and H4) that The number of contributions mediates most of the influence of events on page views, Minor events generate a greater number of page views than other events, and Adverse events generate large increases in page views above those in major or minor events. There needs to be further study to analyze these three hypotheses. Finally we found no support for the hypothesis (H2) Major events generate a greater number of page views than other events.

This research primarily studied the aftermath of events. Some of the questions that remain unanswered by this research are (1) the effect of the different kinds of events on click-throughs; (2) the effect of the event on the number of page views prior to the event; and (3) understanding the effect that different types of visitors to the website have on page views and contributions. These questions will be investigated in future research.

Chapitre 4

Online Community User Influence: A Study Using User Status

Abstract

This research in progress will try to understand the relationship between user influence and status of users in an online community. For online communities a major issue is how to continue to grow and sustain the viability of the community. Oftentimes online communities will gain popularity and then lose that popularity due to stagnation in the number of new users and loss of interest by current users. This research tries to understand how the status of the user plays in the influence of the high level users and the vibrancy of the online community. We measure the influence of the user by using a pseudo Hirsch index to measure the interest of the users posts and use the status of the user in the online community to gauge the user status in the online community.

Keywords

Hirsch Index, Influence, Online Communities, Replies, Threads, User Status, Views.

4.1. Introduction

How can a website owner continue to maintain the vibrancy of an online community (OC)? Maintaining vibrancy is a constant fundamental question that all owners of successful OC websites have. Once a website has garnered interest and gained a fundamental amount of users, the health and vibrancy of the website is dependent on the use, submission, and interest in the website by the users. However many OC websites tend to die a gradual death as users start to post less and less, and interest in their OC wanes to where activity levels gradually decline to where new material is rarely posted.

For a “fan” community website that does not rely on a tangible product that the website is ‘selling,’ the reliance on material posted by users is more crucial and the understanding of how an OC can continue to be relevant is crucial. The owners of this type of website know the importance of a constant stream of news and opinion posting by the users and the reasons for most users to read the websites. Typically these websites rely on advertising that is directly related to the number of users that visit and read material on the OC website. .

There is a lack of research that looks at measuring user contributions. This paper continues the research on the use of Hirsch statistics (Young 2011a; Young 2011b) to measure the contribution levels by users. We utilize the measures proposed by Young et al. (2011a, 2011b) to see whether high level users, measured by their user status, are really the ones that are influential in the OC. The rest of the paper is organized as follows. First we give an overview of past research, then we explain the research context and the user status levels. Third we explain the measures that we use, and propose how we intend to collect and analyze the data. We finish with a summary and conclusion.

4.2. Literature Review

In this paper we use ‘Online Community’ (OC) to mean “aggregation of individuals or business partners who interact around a shared interest, where interaction is at least partially supported and/or mediated by technology and guided by some protocols or norms” (Porter 2004). In our paper the shared interest is a university sports team. The individuals are the fans of this university who are geographically dispersed around the globe. The technology used is a website created as a means to create a forum for fans of this university and support/mediate their opinions, questions, and answers about their collective interest in the university sport team.

The owner of this website (an author of this paper) has been able to use the growth in interest in the use of the internet and the associated decrease in technology costs to be able to easily create, maintain, and promote the site's OC. Generally, website owners may increase their revenue from advertising by showing an increase in interest and, ultimately, usage of their OC. The literature suggest that OC success is tied to user participation (Assmann et al. 2009; Cothrel and Williams 1999) or a sense of being in a community (Zhang 2010). While user participation is key, garnering high levels or participation does not guarantee the health of the OC over time and user participation is not static (Mousavidin and Goel 2009) and other research suggest that OC's evolve in a distinct community lifecycle (Iriberry and Leroy 2009). The challenge that OC owners face is to what factors about the OC allow the OC to have a long lifecycle or extend the length of the vibrant period of the OC's lifecycle. Also, status is described as a way for businesses to shift through flotsam and jetsam to find potential ideas (Fuller et al. 2004; Gangi et al. 2010).

One gap in the literature is the limited research conducted on small, independently owned OC's with only hundreds or thousands of users. Conversely, much research has been conducted on larger more popular OC's such as facebook, Friendster, myspace, or ebay (Clavio 2008).

Two other gaps are the limited research in understanding what drives page views in a smaller OC and how to measure influential users in an OC. Past research in this stream begun to address this issue (Young 2011a; Young 2011b). We utilize the measures identified this research stream.

The fourth gap is the lack of measures to identify influential users. Young et al. (2011a, 2011b) identified the Hirsch index, a bibliographic measure, as a potential way to be modified and used to identify influential users. This pseudo-Hirsch measure is a way to measure both the number of contributions to an OC and the uptake of those ideas via 'replies' to the posting and 'reads' of the posting. Explanation of the development of the measure is done in the measures section below.

The literature is full of research that is based on user status and how it relates to online community usage. There is research on the evolution of the hierarchy of user status (Stewart 2005), the incentive of posting to OC's (Cheng and Vassileva 2006) and the use of status as an incentive (Cheng and Vassileva 2005), as well as the role of seeking status and gift giving to increase status online (Lampel and Bhalla 2007). This extant research on user status finds that most OC's use a direct measure of frequency of contribution as the methodology to measure the status of the user.

Thus, the bibliometric measure is a better indicator of highly influential users in the OC. We therefore propose the use of the pseudo-Hirsch measure as a better way to identify the user status and will compare this to the current user status of an OC of a college sports fan website.

4.3. Research Context

BigUFans.com is a fan website for Big U and their athletic teams. Within the website is an OC where users can post messages and information to other users of the website. There are five major websites dedicated to Big U but BigUFans.com is the largest free and independently owned community for Big U fans. The site was created in 1999. BigUFans.com's primary source of revenue is banner advertising. BigUFans.com has several third-party ad aggregators (such as Google AdSense and Burst Media) that sell and deliver ads to BigUFans.com. BigUFans.com is owned by an alumnus of Big U. There are 23 moderators and four administrators of the website. Both moderators and administrators evaluate content and monitor the message boards for any spam, duplicate threads, and incendiary posts. When troublesome posts are found the moderators/administrator removes such posts. They can also ban users if warranted. Feedback from users of BigUFans.com tends to be positive on the fact that its message boards tend to stay on topic and are free of flaming, incite inducing, and advertising posts.

Big U is a university in the southeast U.S. The school has athletic teams that compete in NCAA Division I and the major sports are football, basketball, and baseball. The school is very competitive in football and the majority of fans post about the football team in the message boards. During the 2009 football season (August through January) the site averaged 2,094,226 page views, 2,219 new thread topics, and 151,286 unique visitors per month. The average visitor viewed 6.6 pages per visit. On an average day the site was visited by 1,075 registered users and added 13 new members. The demographics are 92% male, college educated (72%), and middle to upper income (71%). During the football off-season the site averages 30,000 – 40,000 page views per day. During this time (April 9, 2010 through June 14, 2010), of the signed in users, less than 4% created a new thread and less than 30% posted a reply.

Currently BigUFans.com gives status to users dependent on their posting activity. Anyone that creates an account and signs in is considered to be a 'citizen' of the OC. But depending on the amount of activity (in the parenthesis) a user can reach the following status: BigU Scout Team (0 or more), BigU Third Team (13), BigU Second Team (50), BigU First Team (250), BigU All-Conference (1000), BigU All-American (2000), BigU Hall-of-Fame (5000). The breakdown of users is shown in table 4-1. To gain status on the BigUFans.com a user must sign in and then create a thread or post a reply to a thread to get their posting counted towards a higher status.

Table 4-1. User Status Breakdown

User Status	Count
Scout Team	20363
Third Team	1348
Second Team	1086
First Team	586
All-Conference	173
All American	118
Hall of Fame	44
Administrator	5
Moderator	13
Banned	856

4.4. Measures

These are two main categories of variables needed for our measures: Message Board and User Variables.

Message Board Variables. A 'post' is the first contribution or any reply within a thread. A 'thread' is a collection of posts including the first contribution and replies to that contribution. A 'thread' is the seed to a discussion that garnished one or more replies. An individual 'post' can be a 'thread' when it is the first 'post' without a reply. The reply is a 'post,' as well, but is not a 'thread'. Thread id is used to uniquely identify a specific 'thread'. Reply count is the number of replies to a 'post'. Thread 'views' is the number of views that each thread receives.

User Information Variables. Post 'user id' is the unique identification of a user that created a post. 'Threads created' is the number of threads created by an individual user. 'User status' is the status of the user at the time that the data is pulled from BigUFans.com

A simple count of postings is not particularly useful as people may post 'fluff' posts in order to gain status, or 'fluff' posts in general are counted towards the user status regardless of intent. While posts do create a notion of vibrancy, the real postings that have influence are the ones that garner views by other users. Thus, these influential posts gain responses and create lively discussion. We use the Hirsch index to measure the amount of user influence..

In bibliometrics the Hirsch index (h-index) utilized citation counts of all the publications by a subject to measure the impact of that subject on the field. The h-index measures both the amount of work that a subject has published as well as the impact of that body of work using citations by other researchers to measure the 'influence' of the subject. Formally the h-index is defined as: "A scientist

has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each.” (Hirsch 2005)

The h -index is responsible for creating a buzz in the bibliometric field spawning other measures such as the g -index (Egghe 2006) and the hc -index (Sidiropoulos et al. 2006). While there is debate as to the validity and usefulness of the h -index and other h -family of indices, areas such as information systems, chemistry, physics, and economics have used this measure to show influence of the researcher (Glanzel 2006; Truex et al. 2009). The h -index is becoming accepted in academics as a measure of influence.

We find similarities to posting in an OC and academic publishing. First, a user posting is similar to a publication by a researcher. For the OC we have two measures that are similar to the citation of a paper. We are able to measure the ‘replies’ similar to the citation and we are also able to measure the ‘views’ or the posting being read by an individual user. The ability to measure the ‘views’ is an advantage of the OC that we do not see in bibliometrics. While it is impossible to see if a researcher has ‘read’ another researchers article with the OC we can see if the posting has been ‘viewed’ or in essence opened to be read. So we are able to find two types of h -indices for the OC.

We termed the two new measures as the hr -index (replies) and the hv -index (views). To compute the hr -index one must first take each user in the OC and rank the user’s threads started by the number of replies each thread has garnered. The user has an hr -index of hr where they have at least hr threads with hr replies and the remaining threads have $hr-1$ replies or less. For example, if a user has an hr -index of 100 then that means the user has started at least 100 threads that have gained at least 100 replies or more each. So this is an index of both productivity and impact via replies. The process for obtaining the hv -index is the same except the thread are sorted and counted using views instead of replies.

For each of the indices we use Google analytics and the site’s message board database to export the data to excel and list by the thread id, authors user id, user status, the number of replies, and the number of views to the thread. The data will be then fed into a Java program to rank and list the replies and views. The two lists will then be analyzed to create an hr -index and hv -index for every user that has created at least one thread.

4.5. Research Proposal

We propose to take hr and hv -index data from the BigUFans.com website and compare this to the user status. We hope to compare the hr , hv -index numbers with the status to see if there is a

correlation between status and influential postings. We suspect that there will be some correlation between status and high hr, hv-index numbers because one must post a number of threads to gain a high hr or hv-index but that the correlation will not be that high as there will be many users with high status that have little influence on BigUFans.com in general.

We realize that the hr and hv-indices will be magnitudes higher than the h-index that a researcher might see with their publications. This is due to the fact that publications are much harder to publish than thread postings, and citations take much longer to garner compared to 'replies' and 'views' on an OC.

4.6. Summary

This paper extends past research by Young et al. (2011a, 2011b) as well as research on user status and their contributions in OC's. Two pseudo h-indices the hr and hv-indices were used to measure the influence of the user. We propose to compare the influence measures with actual user status of the posters to see if their status matches up with their influence measures. We expect some correlation due to the fact that in order to gain hr or hv-index numbers one must post and therefore posting leads to higher status in the OC. But we also expect that highly influential posts to come from posters with lower status.

There is a risk in translating the h-index to OC's. First the OC hr and hv-indices are bond to be higher in numbers compared to the bibliometric h-index. This is due to the fact that is more difficult to publish in academics and getting cited by others in academic publishing. While the hr and hv-indices are tied to thread initiation, replies and views to these threads, which is much easier to gain in a shorter period.

There are also a couple of limitations to trying to translate the h-index to OC's. The views are just clicks on a thread and downloading the thread to a computer screen, and are not necessarily a reading of the thread. Second the reply may be a genuine reply to the thread or may be off topic to the thread. We believe that there are two reasons to overlook this limitation. First, the administrators of BigUFans.com are dedicated to the integrity of the message board and minimize the replies to run off topic. Second, in order to assess each reply one mush sit and read all threads and replies to see if they are indeed replies to the original thread. This is not in the spirit of the creation of the h-index. The h-index was created to allow a quick assessment of influence and to avoid having to read all articles published by an author. In the same notion we use the hr and hv-

index to allow us to measure influence without reading all threads. We believe that the hr and hv-indices should be applied in the same spirit as the h-index.

We realize that this research was conducted on a college sports fan site and may not be generalized to other types of OC's. But we believe the contribution of the use of hr and hv-indices to OC's and as a measure of user status can be generalized to other types of OC's. We also note that this was done on a college fan OC and not a traditional business website. We argue that with the competition of college fan websites and the revenue dollars generated this type of OC is a legitimate business site and worthy of research.

Another problem with the current application is that users that have been on the website longer have the ability to gain higher status and hr, hv-indices. A similar second problem is that there might be users that contributed little up to this point but created threads that have garnered much discussion. The second problem is the 'one hit wonder' problem where there is a problem in how to assess the user who has created a thread that spawned a discussion with many replies (one hit). These are two common problems with the h-index and were dealt with the g-index (Egghe 2006) that addressed the time issue and hc-index (Sidiropoulos et al. 2006) which addressed the one-hit wonder issue. Perhaps future studies can look at a comparable g-index and hc-index for OC's and their application but we feel for this study a progressive step of using the hr and hv-index to measure status was sufficient.

Some additional questions that future research may be able to address are (1) What is the effect of specific events with BigU on user participation? (2) What is the effect of events with BigU on the number of page views (replies) prior to and after the event? (3) How are revenue and the effect of different types of users to visit the website related? (4) How does the users hr and hv-indices change over time? (5) can we generalize our findings to OC's of different types?

We feel that this research is relevant to both the user and developer of OC's. First the research showed that most OC's use a simple count for user status. While this is the defacto method at this point we feel that developers of OC's would be interested in the actual influence of user participation in an OC. From the user point of view a better measure of user status would better assess the posting authors and allow user to use the user status measure more accurately.

4.7. Conclusion

This research is part of a overall research stream that is trying to find out how developed OC's can continue to grow and generate page views and interest of the community and stay viable in the eyes of the users. The current study tried to address the issue of OC's from the user point of view. In

particular the study addressed the issue of user status and whether users with high status are truly influential to the OC. User status is relevant to the users as user status is an important indicator of how well you are contributing to the OC. Owners of OC's would be interested in this research to find out whether a simple count of contributions really is a good measure of influence.

Chapitre 5

What Drives User Contribution in an Online Community ? A Study in Contributor Influence and User Status

Abstract

Online communities (OC's) depend on shared interests and user interactions mediated by technology. Successful OC's find ways to encourage these interactions to grow communities. Many OC's have influential users that help grow the community by their very presence and contributions. However, the process for identifying users having the greatest impact is not trivial. This study offers a new method for identifying these influential users through the creation of modified Hirsch indices, which improves upon the current method of using contribution counts or a survey method of polling other users. We validate the new measures against user status and then analyze the measures by correlating them against postings, thread starts, and views and replies to the thread starts for a shared interest OC.

Keywords

Hirsch Index, Influence, Online Community, Online Identity, Replies, Threads, User Status, Views, Virtual Community.

5.1. Introduction

An online community (OC), sometimes called a virtual community, is defined as being an “aggregation of individuals or business partners who interact around a shared interest, where interaction is at least partially supported and/or mediated by technology and guided by some protocols or norms” (Porter 2004). Our research looks at OC’s from the point of view of the health of the OC and how to maintain vibrancy in the OC. Unlike a website that is ‘selling’ a tangible product, owners of an OC are offering a product requiring interaction and visits from the OC’s users. The OC is dependent upon these interactions and visits to remain healthy and vibrant. As long as the users of the OC consider this service to be of value the owners can reap the benefits of providing this service. From the point of view of the users, the OC is of value when the user is able to find relevant and interesting information on the OC. But the producers of the service of the OC are also the users, so the users are both the producers and consumers of the content of the service. So the success of the OC is in the interest of both the users and OC owners.

Successful OC owners are interested in maintaining OC vibrancy. Once an OC has gained interest and a critical mass of users, the vibrancy and health of the OC is dependent on the production of the content via user submissions, viewing and reading of this content, and interest in the content by the users of the OC. Many OC’s struggle to maintain this vibrancy as interest, use, and contributions start to decline signaling the death of the OC.

This study examines a niche ‘fan’ community OC. The OC provides a service of informing the community about the communal interest in a large university’s athletics program. The owners of this type of OC realize that user contributions are crucial to the survival of the OC and that user-generated content is why other members of the community visit the OC and read the postings, and thus, positively contribute to the health and growth of the OC. Typically these OC’s rely on banner advertising which generate revenue only when many users visit the site and see these advertisements. Thus, user vibrancy is of utmost importance in these contexts.

This paper examines the effect of user status on the vibrancy of the OC. Based on social presence and social influence theory, some have argued that contributions to a web site by high status users will increase the vibrancy of the OC and thus contribute to the success of the site (Assmann et al. 2009; Cheng et al. 2006). To examine this idea, this paper continues the line of research that uses bibliometric indices to assess the influence of contributors to the OC (Young 2011a; Young 2011b).

The rest of the paper is formatted as follows. First, we introduce some background research in the area of OC's. Second, we introduce the context of the research. Third, we explain the methodology used. Fourth, we present the data and data analysis. Finally, we conclude with a discussion of our findings.

5.2. Background

The success of an OC is tied directly to user participation. An OC is only as good as the vibrant contributions made by the users of the community (Assmann et al. 2009; Cothrel et al. 1999). OC's that die out tend to lose interactions from the users and the material becomes outdated and communication in the OC becomes stagnant. Even with high participation by the users, this does not guarantee the health of the OC as user participation is not static (Mousavidin et al. 2009). OC's tend to evolve in a distinct community lifecycle (Iriberry et al. 2009), with a beginning period, a middle period where the usage of the OC becomes more and more vibrant, and an ending period where the OC tends to go out of fashion and loses the interest of the OC community. The challenge for OC owners is to extend the middle growth period, or at least maintain the exchange level during the middle period, as long as possible so that the OC participants do not lose its interest in the OC. Examination of smaller OC's is lacking (Shen et al. 2006), however, many papers have looked at larger OC's such as MySpace, Facebook, and eBay. The smaller OC's have been mostly ignored until recently (Clavio 2008).

5.2.1. Assessing OC's based on User Status

One way to analyze the use of OC's is to look at the users motivations and interests. OC's are sometimes a place where users are able to find a sense of community (Zhang 2010). Many OC's create a sense of community by having a common cause or goal for the users. Many OC's also incorporate a hierarchical user status (Chen et al. 2011; Spring et al. 2008; Stewart 2005), which can be used as an incentive to induce user participation (Cheng et al. 2006), but also create users whose sole goal is to gain user status (Lampel et al. 2007).

User status research has focused on social capital (Ganley et al. 2006; Law et al. 2008), social presence (Shen et al. 2009; Shen et al. 2006), and social identification (Shen et al. 2008; Zheng et al. 2007). According to social capital theory, users are more likely to exchange intellectual capital and create new intellectual capital when structural, cognitive, and relational dimensions are met (Nahapiet et al. 1998). The structural dimension is where users are connected through a network of ties. The cognitive dimension is where the users share a narrative and language. The relational

dimension is where users have strong relational characteristics such as trust, norms of cooperation, obligation, and identification. One dictionary definition of identity is “the individual characteristics by which a person or thing is recognized” (Collins 2010). User status ties to this notion of identity or identification.

In social presence theory, presence is defined as “the perceptual illusion of nonmediation” (Lombard et al. 1997). The term ‘perceptual’ means “this phenomenon involves continuous (real time) responses of the human sensory, cognitive, and affective processing systems to objects and entities in a person’s environment” (Lombard et al. 1997). The ‘illusion of nonmediation’ happens “when a person fails to perceive or acknowledge the existence of a medium in his/her communication environment and responds as he/she would if the medium were not there” (Lombard et al. 1997). This theory suggests a substantial mental model of a user “is activated immediately upon detection of behavior that suggests the presence of another intelligence”, where the social interaction is mediated by technology (Biocca et al. 2003). Finally social identification theory predicts a user’s identification with an OC “that one has chosen volitionally stems from an understanding that membership entails significant benefits” (Dholakia et al. 2004).

5.2.2. Assessing Vibrancy of the OC

Another issue in OC research that targets smaller OC’s is the reliance on survey research, administered either to users or owners of the OC (Law et al. 2008; Marett et al. 2009; Schaedel et al. 2010; Shen et al. 2008; Shen et al. 2009; Zheng et al. 2007). These studies utilize survey tools and then depend on the ability of the survey taker to measure the validity of contribution of users in the OC. Problems with this methodology exist, however. First the data collected is susceptible to the survey takers opinions and feelings, which are susceptible to biases of the survey taker. Another problem is that the answers to surveys and the instruments are inexact. Some ‘wiggle room’ exists with the use of Likert scales, both on the part of the survey takers and the researchers whom interpret the data. We propose a more suitable source of information would be based on observation of the behavior of the users in the OC. We develop a measure that is based on the actions of the users in the OC. In particular we are targeting the ‘views’ of postings and the ‘replies’ to posting that are initiated by the OC users. We describe this measure, which is based on the Hirsch index, later in the Methodology section.

This paper expands current research on the use of the Hirsch index (Cuellar et al. 2012) to measure the contribution levels of the users. The Hirsch index (h-index) utilizes the citation counts of a publication in order to measure the impact of the author of that publication. The h-index is defined

as follows: “A scientist has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each” (Hirsch 2005). In order to measure the h -index of a researcher you need to sort all publications by the author by citations to each publication. When the rank number of the publication overtakes the citation number, you have your h -index. What makes the h -index powerful is that it is a measure that is based on publication frequency and publication influence based on citations. In order to increase your h -index by one, you need to publish one more paper with more than h citations. To get a higher h -index, a researcher cannot just publish many papers or just gain many citations on one paper, they must do both.

The h -index has created a buzz in the bibliometric literature with papers that praise the use, extend the use, and criticize the use of the h -index (Egghe 2006; Glanzel 2006; Hirsch 2007; Sidiropoulos et al. 2006). Some of the criticism of the h -index have been that the h -index is time sensitive which was addressed with the hc -index (Sidiropoulos et al. 2006), and that the h -index is not sensitive to publications that have a larger than normal impact, which was addressed with the g -index (Egghe 2006). The introduction of the h -index has spawned much research in non-bibliometric fields where they use the h -index to measure the impact of authors in fields such as chemistry, physics, economics, and the IS field (Truex III et al. 2009; Truex III et al. 2011).

We find there are similarities between an OC and academic publishing. We also find the essence of the h -index can be applied to measure impact a user has with their postings in an OC. A user post in a forum is similar to a publication by a researcher. The reply to a posting by another user is similar to a citation of a publication. Applying this h -index based measure will allow us to identify influential users better than simple post counts or user opinion surveys, which are what most OC's are currently using to identify influential users (Chen et al. 2011; Ganley et al. 2006; Spring et al. 2008).

5.3. Theoretical Framework and Hypothesis Development

User status is one of the key drivers used by OC users to determine the importance of postings. This conclusion is deduced from the relation dimension of social capital theory and social presence theory. First, from the relation dimension we apply the concept of identification, where a user in an OC is recognized in some fashion. We use user status as identity, which is a public measure that is recognized by another user on the OC as a measure of the social worth of the first user. When a user goes on to the OC and browses through the postings they will attach more validity and trust to the postings by users with relatively higher status. According to Social Presence Theory, once a user starts to peruse the postings, the user's mental model will be activated and more likely to click on a posting when the user detects or identifies the author of the posting to be of a higher status or

'another intelligent' user. With the OC we are analyzing, we are able to see how many times a particular posting is viewed and replied to. The use of the posting in the community can be measured by the number of times that the users read and reply to threads. With the data we have gathered we will be able measure the user status by using the OC assigned user status, the modified h-indices for a time period and see if these correlate with future views and replies. Thus we have a set of hypotheses:

H1a: Contributions from users with higher user status will have an increased number of views in the OC.

H1b: Contributions from users with higher user status will have an increased number of replies in the OC.

We posit that users with higher user status have already chosen the OC, which means they have chosen the OC 'volitionally' by understanding the benefits of joining the OC. From Social Identification Theory and Social Capital Theory, we learn that once a user is vested into using the OC they will tend to keep contributing to the OC, hence continuing to create more intellectual capital and reaping the benefits of the OC. So, users who gain users status will also be more likely to start threads and continue to create postings. Thus,

H2a: Users with higher user status will start more to threads on the OC.

H2b: Users with higher user status will contribute more posts on the OC.

Since traditional measures of user status are based on the number of postings made, our H2b may seem like a tautology; however, this is not the case. We are looking at higher user status in one time period and the thread starts and postings in a subsequent time period. We split our data collection process into two distinct phases. First, data on posts was collected for one year to establish the traditional and h-index based measures of user status. Next, another year of data was collected to test the hypotheses. User status is held constant at the level achieved through the first data collection period, and the posts used to determine the measures are not included in the data used to test the hypotheses. This ensures posts are used once, either to determine the measures or to test predictions, never for both.

5.4. Research Context

BigUFans.com is a website for the athletic teams that represent Big U. Big U is a public university in the state university system, over 150 years old, and located in the southeast United States. Big U is a

large tier-one research university with over 30,000 students. The major sport for Big U is college football, although they field competitive teams in basketball and baseball, all of whom compete in the NCAA Division I.

There are five major OC's dedicated to Big U. BigUFans.com is the largest free and independently owned OC of these five. BigUFans.com was created in 1999 and is owned by an alumnus of Big U, who is one of the authors of this paper. BigUFans.com allows anyone to browse postings and content. However, a user must be registered and signed into the site before creating original posts or replying to original posts. The website is completely free to its members and visitors and generates almost all revenue from advertising. The health and vibrancy of the BigUFans.com website as measured by site traffic, user contributions and interactions is crucial for its owners.

There are 17 moderators and six administrators of BigUFans.com. Administrators are responsible for all aspects of running and monitoring the community. Moderators are given specific topic boards within the site to monitor and moderate. Registered members may become moderators or administrators only at the invitation of the publisher of the site. Both administrators and moderators are expected to evaluate the content of message board postings for possible spam, duplicate posts, posts in wrong topic folders, advertising, and incendiary posts. When any of these are found administrators and moderators are able to remove such posts or place them into the appropriate areas, as well as flag the users who created the posts for possible banning of their user accounts. Feedback from the OC community tends to be positive in that the message posts normally have less flaming, incendiary posts, advertising, and off topic postings.

During the 2011 football season (August through January), BigUFans.com averaged 2.9 million page views, 1075 new posts, 89 topics started, and 167,000 unique visitors per month. During the season, on average, the OC is visited by 1,017 signed in users and adds 5 new members per day. On a typical visit the average user views 5.2 pages. The users on average consist of males (92%), hold a college degree (72%), and are of middle to upper income (71%). During the 2010 off-season (February through July), BigUFans.com averaged 2 million page views per month. During 2011 less than 4% of the signed in users created new threads and less than 30% posted a reply. Currently, BigUFans.com automatically gives user status to members according to the contributions as measured by the total number of posts that the user has made in the message board (see table 5-1). This status is displayed under each user's username for all postings in BigUFans.com message board.

Table 5-1. BigUFans.com user status/posting required and coding of StatusNo.

User Status	Number of Posts Required	Number of Registered Users with this Status	StatusNo
-------------	--------------------------	---	----------

Administrator	n/a	6	n/a
Moderator	n/a	17	n/a
Scout Team	0	20470	1
Third Team	13	1363	2
Second Team	50	1086	3
First Team	250	590	4
All-Conference	1000	179	5
All-American	2000	121	6
Hall-of-Fame	5000	47	7

5.5 Methodology

Data was collected on BigUFans.com for calendar years of 2009, 2010, and 2011. This data resulted in 3632 distinct authors and 70,825 distinct threads started. Data for each thread includes a user identification (userID), the number of views that the thread received, and the number of replies to the thread. This allows us to identify the impact of the thread started by each user. We also tied each userID to their user status, number of threads that they started, and the number of postings that they authored. Postings include both thread starts and replies to another user's thread. We tied numbers to the user status: Scout Team (1), Third Team (2), Second Team (3), First Team (4), All-Conference (5), All-American (6), and Hall-of-Fame (7).

Hirsch indices were calculated for each userID after data was imported into a spreadsheet. A Java program was written to allow the calculation of the h-indices. This allowed us to measure the impact of each user. We then imported the resulting impact measure data into SPSS to calculate correlation values and evaluate the hypotheses.

5.5.1. Measures

There are two primary variable categories used in this study: thread views/replies and user posts. Two distinct measures can be associated with a thread: the number of 'views' of the thread and the number of 'replies' to the thread. While a 'view' constitutes evidence that a thread was opened by someone visiting the OC and the thread was shown on the screen to the visitor, a 'view' does not necessarily indicate that the visitor has read the posting. A 'reply' on the other hand is a visible cognitive response to the thread posting whereby we can assume that there is some level of reading by the responder, and thus, some cognitive influence by the thread starter on the responder. This assumption is strengthened by the moderators culling off topic posts and "spam" posts, which would not indicate any influence of the original post on the replies.

Two types of h-indices measure the impact. First, we measured the impact of each thread started by a unique user to find the user influence on other OC members. Second, we identified the hv-index which is calculated similar to the bibliometric version of the h-index (Egghe 2006). Third, we selected all threads started by a unique userID and ranked them according to the 'views' to each posting. The point when the rank number was smaller than the actual 'views' was identified as the hv-index. The hr-index is calculated in the same manner, except the ranking was calculated by using 'replies' instead of 'views'.

The second category of variables includes the number of postings associated with a particular user. Postings were measured by associating each unique userID with the number of threads started and the number of total postings by that userID. The 'threads started' are the set of postings associated to the userID's that initiated each thread. This means that for each thread we identified, the author of the first posting was credited as having started a thread. The number of 'total postings' is the total number of postings that a particular user made to the message board, including both thread starts and replies to threads. For example, if a user has an hv-index of 100 that means that the user has at least 100 thread starts that have garnered at least 100 views each. If a user has an hr-index of 50 that means that the user has at least 50 thread starts that have garnered at least 50 replies each.

5.5.2. Instrument Validation

To validate the two new measures, hr and hv, we compared them to the old measures of user status. First, we calculated the measures for each user id, using one calendar year of data (2009 and 2010) to initialize the measures. Second, we used the 2010 and 2011 calendar-year data to observe the views and replies of thread starts by each userID. Third, we ran correlation analysis between the user status, hr, and hv measures for 2009 and 2010 and the reply and view numbers for 2010 and 2011, to see how well each of the measures is able to predict user actions. The output is measured by the threads started, posts submitted, replies to posts, and views to posts, all measured for distinct user ids.

The data was run in SPSS and the correlation matrix using Kendall's Tau is presented in Table 5-2. The status number (Status) was assigned according to Table 5-1. The highly correlated predictor for each set of years is indicated in **bold**. We have data for two sets of consecutive years (2009 to 2010, and 2010 to 2011) and data from two years separated by one calendar year (2009 to 2011). For 2009 to 2010 (lower right quadrant of table 5-2), we see that hv (in 2009) has the best predictive power as hv correlates best with threads, replies, and views to these threads (in 2010). Status is second and the hr-index is weakest to predict user impact in 2010, only achieving about half of what hv is able to

predict. For 2010 to 2011 (upper left quadrant of table 5-2) we see that still hv is top in three of the four output measures. For both years we do see a spike in the prediction power of status to predict user posts in the subsequent year. This is not too surprising as user status is a direct measure of posts.

We have on dataset across two years with predictor measures made in 2009 to predict output in 2011 (upper right quadrant). Note that over a period of two years user status has lost it's predictive power, where even the hr measure is a better predictor of user behavior and impact two years in the future. Again hv is the best at predicting the four output measures across two years. So, while we need to expand this dataset to include the lifetime of the OC, we cautiously conclude that the hv measure seems to be the best predictor, and hr is a weaker predictor, of future user behavior and impact. Note that the lower left quadrant of data was included for completeness but is data form 2010 predicting behavior in 2010, which is tautological and thus the data here is highly correlated.

Table 5-2. Correlation Table of Predictor variables vs. output measures using Kendall's Tau

		2010 hr	2010 hv	2010 Status	2009 hr	2009 hv	2009 Status
2011 Threads	Correlation Coefficient	.089(**)	.253(**)	.242(**)	.046(**)	.147(**)	.005
	Sig. (2-tailed)	.000	.000	.000	.002	.000	.746
2011 Posts	Correlation Coefficient	.144(**)	.288(**)	.457(**)	.082(**)	.179(**)	.171(**)
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
2011 Replies	Correlation Coefficient	.089(**)	.241(**)	.233(**)	.048(**)	.140(**)	.011
	Sig. (2-tailed)	.000	.000	.000	.001	.000	.425
2011 Views	Correlation Coefficient	.084(**)	.239(**)	.229(**)	.044(**)	.136(**)	-.002
	Sig. (2-tailed)	.000	.000	.000	.002	.000	.889
2010 Posts	Correlation Coefficient	.425(**)	.637(**)	.933(**)	.179(**)	.302(**)	.411(**)
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

2010 Threads	Correlation Coefficient	.641(**)	.971(**)	.689(**)	.151(**)	.301(**)	.224(**)
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
2010 Replies	Correlation Coefficient	.659(**)	.865(**)	.655(**)	.159(**)	.300(**)	.232(**)
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000
2010 Views	Correlation Coefficient	.632(**)	.900(**)	.666(**)	.151(**)	.293(**)	.220(**)
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

** Correlation is significant at the 0.01 level (2-tailed).

5.6. Findings and Hypotheses

From table 5-2 we find that hr, hv, and status were positively correlated with views (hr, hv, and status respectively, .084, .239, .229, .151, .293, .220 for one year, .044, .136 for two years) and replies (.089, .241, .233, .159, .300, .232, for one year and .048, .140 for two years), which confirmed H1a and H1b. For one year, status is good at predicting views and replies, but we found that over two years the h-indices were better at predicting views and replies. In fact, in the two-year data, status was not able to predict any user behavior or impact except for postings. We conclude that the h indices can be used to predict user impact over multiple years while user status gives us indication only over the subsequent year.

The three measures were also positively correlated to threads (hr, hv, and status respectively, .089, .253, .242, .151, .301, .224 for one year and .046 and .147 for two years) and posts (.144, .288, .457, .179, .302, .411, for one year and .082, .179, .171 for two years) confirming H2a and H2b. Here we see that status is actually a good predictor over two years of posts and a strong predictor of posts over a one-year period. This is not surprising as status is gained through posts. The drop off of users that come and go explains the loss in predictive power over a longer period of time.

5.7. Summary

This research extends past research by (Cuellar et al. 2012) as well as the use of user status in OC's. Two variants of the h-index, the hr and hv-indices, were used to measure user influence. We set out to see if user status really did correlate with the users influence in the OC.

There are several risks for taking the h-index and using it in OC research. First, the hr and hv indices are bound to be higher than h-indices in bibliometrics. The bibliometric h-index is tied to citations of publications while the hr and hv-indices are tied to views and replies in a message board system. Publications in academia are more rigorous outputs that take a much longer time to develop than a posting on an OC. Given this, we expect the frequency and the raw numbers to be higher for the hr and hv-indices compared to the h-index in bibliometrics. Second, with the hv-index, a 'view' is just a click on a thread and the subsequent download and showing on the computer monitor. A 'view' does not necessarily indicate 'reading' by the user.

Third, with the hr-index, a reply may be related to the topic at hand but may also be completely off topic. A reason for an off-topic reply may be that the user is trying to gain higher status without actually making a valid point in their reply. We believe there are two reasons to overlook this limitation. First the administrators and moderators try to make sure such off-topic replies are dealt with and deleted. Second, in order to assess the validity of a reply one must read all replies to all posts. Reading everything is not in the spirit of the h-index, which was created to allow a quick read of influence without having to read all articles that are cited.

Fourth is a problem inherent in the h-index. The length of time that a researcher has been publishing has direct impact on the h-index. The longer a researcher has been publishing the more likely the researcher is able to garner a higher h-index. Similarly, the hr and hv-indices are both dependent on the amount of time that the user has been active on the OC. Fifth, also inherent in the h-index, is the researcher that has only a few highly cited papers will not gain a high h-index. This is the 'one-hit-wonder' problem. The hr and hv-indices suffer from this as well. Those users that begin only a few threads that have many responses and a high impact, will be influential in the community, but will not appear so in the indices. In bibliometrics the g-index (Egghe 2006) dealt with the 'one-hit-wonder' problem and the hc-index (Sidiropoulos et al. 2006) dealt with the time problem. Both of these may be extensions to this research stream for OC's in the future.

We also realize the use of this particular college sports fan site may not generalize to all OC's. However, we believe the contribution of using the hr and hv-indices, which improves on the survey methodology, is that the measure can be generalized to assess user contribution in many environments. We also realize the lack of a direct business use by the users of a free sports fan site

but argue that competition in college fan OC's is high and that the revenue dollars generated by this OC constitute a legitimate business.

Some additional possible future research may look into the questions of (1) How specific events involving BigU influence user participation? Specifically how does an event affect page views prior to and after the event? (2) How do the hr and hv-indices of a user change over time? (3) Can our findings be generalized to other OC's and/or others areas in IT (such as knowledge management) involving user participation?

We believe the current research is relevant to academia and the users and owners of OC's. User status is correlated to a simple count of postings for the OC studied. While this count is the only method of granting user status currently, we believe that many OC owners would be interested in a better metric for assessing user contribution and impact in their OC's. If a better user status granting method is implemented, the users would also benefit from increased efficiency when searching through messages to discover relevant postings.

5.8. Conclusion

This research is part of an ongoing research stream that is examining how developed OC's can continue to grow and generate page views, interest, and online social interactions within the community and stay valuable in the eyes of the users. We were able to show the new measures, hr and hv, were better predictors than user status over a period of time. Furthermore we were able to show that the predictors, hr, hv, and user status is correlated with thread starts and postings, and more importantly, is correlated to the impact measures replies and views.

The current research attempted to address an issue faced by OC's from the user's point of view, examining user status. User status is important to a user because it is a measure of how well the user is contributing to the OC and can be used to compare to other users on the OC. We believe that the implementation of a more accurate measure of impact such as the hr and hv-indices can improve on traditional user status measures. We also believe that for researchers, the current use of survey based data collection can be improved by the collection of usage data. This data can be used to calculate hr and hv-indices, which are quantitative measures of impact that can improve on current measures of user impact.

Chapitre 6

Online Community Influence: A Study Using the Hirsch Metric and Social Network Analysis

Abstract

This study looks at small to medium sized online community (OC) and tries to identify ways to measure impact of the contributions of the users of the OC. OC's are dependent on contributions of their users to maintain the health of the OC. Measuring the health of an OC by identifying those users that have most influence and thus create more activity and finally more people the visit the OC is an important activity to the stakeholders in the OC. In order to measure these high end users we are extending previous research to include a two part measure. First, using the Hirsch metrics to measure the productivity and impact of user contributions and second, using social network analysis to see those users that have high centrality measures in the network of posters and readers of the OC. This study looks at one University sports fan site to measure the influence of their users and found some correlation between the Hirsch measures and the centrality measures.

Keywords

Hirsch Index, Influence, Online Community, Online Identity, Replies, Threads, User Status, Views, Virtual Community.

6.1. Introduction

Online communities have become a driving force in many shared interest websites. The OC or sometimes referred to as the virtual community, is defined as an “aggregation of individuals or business partners who interact around a shared interest, where interaction is at least partially supported and/or mediated by technology and guided by some protocols or norms” (Porter, 2004). For OC’s the health of the OC is measured not by sales of a product but by the amount of visitors, which in turn is dependent on the amount of relevant and new content in the OC. While some OC’s may have content that is primarily developed by paid staff writers, many OC’s are dependent on the users for content. One of the problems with OC’s is they tend to fall out of fashion after a lively, vibrant period. For owners of these OC’s, trying to keep the vibrancy period continuous remains a constant challenge. Our research has been concerned with trying to find ways for owners of OC’s to identify how to maintain the health of an OC.

Our previous research has identified the Hirsch index (h-index) as a possible way to measure the ‘high level’ contributor (Cuellar et al. 2012). The h-index is an important measure that takes into account the author of a research paper and measures both the productivity, in the number of publications of the author, and the influence of the author, using the number of citations their publications garner (Hirsch 2005). There has been a plethora of research on the subject of the h-index and it’s validity, which seems to confirm that the h-index is a valid measure of influence. While the h-index seems to measure influence the consensus is that with any measure the h-index still has some exceptional cases in which the measure tends to breakdown. For example those researchers that publish little but publish only a few highly influential papers, such as Einstien, have a low h-index. In order to combat this problem, the h-index should be used as one measure out of a family of measures to paint the whole picture of a researcher. As a result, spin-off metrics such as the g-index, gc-index, and hc-indices have been created. There has also been many studies in the business academic publishing area using the h-index.

While the use of the Hirsch index is an important step to integrate a bibliometric measure that takes contribution and influence as a measure, in the same sense that the h-index alone does not measure the true researcher, we believe the h-index alone cannot be used to measure the ‘high-level’ contributor. We still need to continue to improve on this measure of influence by the individual user of the OC. In this research study we propose to add social network analysis (SNA) in order to improve on the measure of the ‘high-level’ contributor.

Another area of measure of a community can come from SNA. SNA has been used in several past IS research and while using 'digital trace data' is somewhat new in the IS arena (Howison et al. 2011), we feel that using the postings to determine the social network is valid for this exploratory research. As a step towards creating and improving an aggregate measure of influence of a contribution to an OC we are adding a SNA to our measure. We are proposing the creation of the social network and the analysis using centrality measures to improve on our aggregate measure of the influence of users in an OC.

Our overall research question is: How can an OC continue to foster growth and relevance? In order to find the answer to this question we are first tasked to find those individuals that contribute to the OC that improve the health and interaction of the OC members. So our current research objective is to find a measure or measures that allow us to identify these individuals who stir up interaction in the OC. One caveat in this stream is that we have not started to address and purposefully so the questions of measuring 'quality'. We are merely looking at influence and the ability of an OC member to cause interaction.

This study examines a university sports-fan community OC. The OC is comprised of users who have interest in the university sports teams and their success. Since this OC's scope is not global in nature and is localized to one geographic region in the US, we find that this is at best a medium sized OC. The OC provides a forum for users to share information, debate, and comment on each other's postings. The owners of this OC understand the importance of the member's contributions because the majority of users come to the OC to socialize and exchange ideas and thoughts about their beloved team. The OC, typical to those of this type of OC, rely heavily on users traffic to the OC in order to generate revenue through the visibility and use of their advertisers ad and links.

The rest of the paper is structured as follows. We look at background information including past studies on OC, the use of h-indices in OC user influence measurement, and social network analysis. We then look at our research question and hypotheses. We then look at the OC that we used in this study and some characteristics of the OC. We then present the data collection process and results. We finish the paper with a summary, including weaknesses and possible future studies, and conclusions.

6.2. Background

OC's are vibrant due to contributions from members of the OC. An OC's health is directly tied to the contributions made by the community of users (Assmann et al. 2009; Cothrel et al. 1999). OC's tend

to die out as they start to have a stagnant user community. There tends to be a distinct life-cycle of OC's where they have a word-of-mouth maturation period of gaining relevance and improved activity, a saturation point where activity plateaus out, and finally a declining period where the OC's tends to fall out of relevancy (Iriberry et al. 2009). The challenge for OC owners is to continue the initial growth period as long as possible and try to avoid the declining period. A healthy OC will continue to grow or at the very least be relevant and have continued postings and interactions. While there is many research on popular very large social networking sites, which have national or global scope, there has been a lack of work done on smaller OC's (Shen et al. 2006) this paper is one such study.

6.2.1. Assessing OC's based on User Status

We have found there has been a lack of studies to identify how to assess the health of an OC. We feel one way to study the health of the OC is by looking at the contribution levels of users. OC's typically create a status level for their users and this is used to indicate how much a user has contributed to the OC. There have been some studies that have looked into this user status (Chen et al. 2011; Spring et al. 2008; Stewart 2005) and how user status can be used as an incentive to participation in the OC (Cheng et al. 2006). User status can be seen as a direct quantitative measure of the user's contribution if levels are tied to the amount of contribution users make to the OC.

6.2.2. Assessing Vibrancy of the OC

We have found many studies on OC tend to rely on more indirect 'opinion' measures using surveys (Law et al. 2008; Marett et al. 2009; Schaedel et al. 2010; Shen et al. 2008; Shen et al. 2009; Zheng et al. 2007). Most OC's have a ranking system where the user gets some level when they achieve a certain number of posts. This is basically a posting count which we feel is not a true measure of who is influencing others. The survey studies typically focus on smaller OC's and target the users and/or owners of the OC. While these studies are valid, they still suffer from the fact that surveys are reliant on the expertise of the users. Thus, they are not an exact measure of the status of the current OC. The data collected is dependent on the opinions, feelings, and bias of the survey respondents. Since surveys typically use Likert scale like measures, there is 'wiggle' room for the respondents and the responses are not exact. They may be thinking their feeling is between a 3 and a 4 but since there is no 3.5 they are forced to choose one or the other.

Instead we use a more exact measure that is dependent on the previous behaviors of the users of the OC. We use a measure that is a pseudo-measure of the Hirsch index (h-index) which uses previous

postings and their popularity by looking at the 'reads' and 'replies' to these postings (Takeda et al. 2012). The h-index is a measure in bibliometrics that is defined as "A scientist has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each" (Hirsch 2005). In order for a researcher to gain a high h-index they must do two things. First they must publish enough research papers to gain a sufficiently high h-index. One can only get a maximum of h equal to the number of publications that they publish. So the h-index is partly a measure of productivity. Second, the researchers publications must garner citations. The second part of the h-index measures the relevance of the researchers publications. In order to get a high h-index, a researcher must publish many papers and those many papers must all get many citations.

The h-index has gained some momentum for use in the bibliometric fields, but has been the subject of some criticism which in turn has produced a cottage of extended measures of the h-indices. There was criticism that the h-index is time sensitive. A researcher requires some time to reach enough publications to even get citations. The h_c -index address this issue with a time aspect giving citations to younger publications more relevance (Sidiropoulos et al. 2006). Another problem was that there were some research papers that garnered unusually large numbers of citations that would essentially get the same treatment as papers with lesser number of citations. Once you get past a certain number, the more citations you get did not matter. So the g-index addressed this issue by adding up all the top research paper citations for a given author (Egghe 2006). We feel that these two issues are somewhat minimized with our target OC. The OC vibrancy is based on current measures of interaction and the data captures interaction up to the last second to when the data is pulled, so there is less of a time issue with the dataset. Also the larger individual posts, while important, do not garner more current interaction. There has been an increase in the use of the h-index in other areas of research outside of bibliometrics such as chemistry, physics, economics, and IS (Truex III et al. 2009; Truex III et al. 2011).

Our past study looked into the use of the h-index and identified this as a valid measure (Takeda et al. 2012). The study identified a major difference between the use of the h-index in the OC versus academic publishing. With OC's there is the ability to see those that replied to a posting and those that read a posting (via the display of the posting on their computer screen).

One important distinction is that we are not making a connection between the users with high h-indices and the health of the OC. While the number of users with high h-indices may not translate directly with the health of the OC, those with high h-indices are creating more users to log into the OC, view posts, and ultimately view advertisements. At this stage of the research stream we want to

identify ways to measure high h-indices users to see if their activity actually turns into a healthy OC. We are trying to find ways for the OC developer to identify those users that have high h-indices so that they can cultivate their talent, or try to encourage users with lower h-indices to create posts that will result in higher h-indices.

6.2.3. Social Network Analysis

Centrality in a social network refers to the notion that the group revolves around a certain individual. If an individual is more central that means there are more people connected to that individual, if they are less central that means there are less connections to that individual. There are several measures of centrality, three which we use are degree, closeness, and betweenness centrality. Degree centrality is a simple count of the connections an individual has with others in the network. Degree refers to the number of edges a node in the network has, or the number of connections that an individual has with the others in the network. In theory an individual can have at most $n-1$ connections when there are n individuals in the social network. This is because the individual cannot have a connection with themselves, but can have connections with everyone else in the social network. When one individual has every possible connection, the distance between that individual and all other persons in the social network is one (Acedo et al. 2006; Eaton et al. 1999; Henry et al. 2007; Otto et al. 2002; Polites et al. 2008; Vidgen et al. 2007).

Closeness centrality is a mean measure of how close an individual is to the rest of the network. The distance between two individuals is measured by number of connections between the two individuals taking the closest connection between the two. For example, if I knew a guy (individual A) who knows a guy (individual B), then the distance between me and A is one and the distance between me and B is two. The closeness centrality takes the minimum distances between an individual and all others in the network and gives the mean of this number. (Polites et al. 2008; Vidgen et al. 2007).

Finally the betweenness centrality is a measure of how an individual is used in the minimum distance connections between any two other individuals in the network. The betweenness centrality of an individual is found by the summing all the shortest distances that go through an individual and dividing by $(n-2)(n-1)$ where n is the number of distinct individuals in the social network (Vidgen et al. 2007).

6.3. Theoretical Framework and hypothesis development

Using the hr and hv indices we were able to identify those users that were influential from the citations point of view. By adding a social network perspective we were able to look at those individuals that are central to the network of communication between the users of the OC or the social network of the OC. While we expect that both h-indices and all three centrality measures that we use to be correlated, we do expect some differences in the strengths. For the h-indices there is much more of a chance of interaction in the social network when someone replies to your posting as opposed to just reading your postings. With a read, there is no chance for the author to actually reply and make a connection to the reader. While when a reader posts a reply, there is acknowledgement of the read and confirmation sent back to the author, which in turn can spark a reply to the reply and thus a connection in the social network. So we expect the hr numbers to correlate more highly with the centrality measures.

For degree centrality, this is a sort of 'who you know'. Closeness centrality is a kind of 'how close is anyone in the network to you'. Finally, for betweenness centrality, this is a measure of 'how well do you know any two people in the network'. We expect that direct connections are made from replies, so we should find that degree centrality to be more closely correlated with the h-indices. This brings us to our set of hypotheses:

H1a: Users that have high replies (hr) will have higher degree centrality in the social network.

H1b: Users that have high replies (hr) will have higher closeness centrality in the social network.

H1c: Users that have high replies (hr) will have higher betweenness centrality in the social network.

H2a: Users that have high views (hv) will have higher degree centrality in the social network.

H2b: Users that have high views (hv) will have higher closeness centrality in the social network.

H2c: Users that have high views (hv) will have higher betweenness centrality in the social network.

6.4. Research context

BigUFans.com is a OC website for the school Big University (one of the authors is an alumni of Big U. and the owner of BigUFans.com). Big U. is a large tier I research institution in the south-eastern US. Big U. is a public state university that has been around for 150 years with over 30,000 students. The OC is in particular geared towards the athletic teams of Big U. where football is the major sport. Big U competes in the NCAA Division I in many athletic disciplines including the major sports of football, basketball, and baseball.

Currently BigUFans.com has four major competitors. BigUFans.com was created in 1999 and is the largest free OC. The OC makes revenue by advertising sales. Thus gaining activity on BigUFans.com is an important activity. The other competitor sites have some aspect of paid areas in their OC. BigUFans.com is the only completely free OC for Big U. Anyone may anonymously browse the contents of BigUFans.com but in order to post to forums a user must register.

There are six administrators and 17 moderators of BigUFans.com that work to keep the topics on the forum of BigUFans.com relevant and free of spam and incendiary comments. Registered members can become moderators as they become more active in the community but only at the invitations of the owner of the site. Because of the work of the staff, one of the major advantages of BigUFans.com over their competitors is that the community feels the OC has more positive messages that are free of advertising, spam, flame wars, and off-topic postings.

The majority of activity on BigUFans.com comes from discussion about the football program. During the 2011-2012 season monthly activity saw an average of 2.9 million page views, 1075 new posts, 167,000 unique visitors, and 89 topics started. Daily activity included 1,017 signed in users and 5 new users. A typical visit consisted of 5.2 page views. The demographics were typically male (92%), college graduate (72%), and of middle to upper income (71%). During the offseason prior to the 2011-2012 season, there were 2 million page views a month. Currently BigUFans.com gives different level user status automatically to users that post a certain amount, regardless of how many users read their messages.

6.5. Methodology

For the measure of the h-indices data was collected over the period that encompassed the football season period of August 2012 until February, 2012. We collected data for the h-indices over a longer period of time than the social network due to the fact that the h-indices take a longer period of time to achieve a spread and to get high h-indices one needs more time for replies and views to become recorded. This data collection resulted in 911 distinct authors and over 500,000 distinct thread creation.

The social network data was taken over a two-month period between December 2012 to January 2013 to give a more snap shot of the social network. Given the amount of data to process for this preliminary study we felt that taking the social network data over a shorter period of time was better. This resulted in over 200,000 reads of the postings by over 1600 unique individuals.

Data for each of the threads includes a user identification (userID), the number of replies the thread received, the userID's of those that replied to the threads, the number of views to the thread, and the userID's of those that viewed the thread. A Java program written by the authors was used to calculate the h-indices and UCINET was used to calculate the centrality measures. The data was then imported into SPSS to calculate the correlation values to evaluate the hypothesis.

6.6. Measures

For data collection in this OC it was possible to find the reads or 'views' of posts as well as the 'replies' to posts. This means there was an additional h-index available compared to the bibliometric version. For a researcher to find out how many people is impossible in the OC we are able to measure how many people clicked on a posting and actually had displayed the post on their computer screen for a 'view'. While a 'view' may not necessarily be a 'read' by the user, we assumed the topic had caused enough interest by the user to prompt the action of actually clicking on the link and displaying the posting on their computer. We call this measure the hv-index

The second aspect of the OC is that there is the possibility of a user to respond or 'reply' to a users post. This is equivalent to the citation in the bibliometric sense. The OC user is prompted by the posting so much that they actually type our a reply to the posting. We found this required more action by the responder compared to the 'view' therefore this number was always lower than the hv-index. We called the h-index of the reply the hr-index.

We also collected the 'view' activity of the users on the OC. We were able to identify the posts and the unique readers of these posts. This allowed us to created the social network of influence in the OC. We were also able to create a directional network as we knew userID of the poster and the userID of the readers. One can argue that the flow of data is going form the posters to the reader. Another argument is that the flow should go in the other direction as the citation of the poster is being received from the reader. We created two networks using both directions for this study. Each is labeled either Rp (for reader to poster) or Pr (for poster to reader)

6.7. Findings and hypotheses

The results of the study included the measure of degree centrality, closeness centrality, and betweenness centrality of both networks. We also found the hr-index and hv-index of each of the posters. We took these measures and ran correlation analysis on the measures in SPSS, the result is in table 6-1. We are looking for correlation between the H numbers and the centrality numbers

(degree, closeness, and betweenness). Some variables have been shortened in the table in order to display the items to be read easier. Also the two different directional networks are listed with Rp (reader to poster) and Pr (poster to reader). First we find the highest correlation between hr and hv, which is expected as those that get more replies are bound to be getting high number of views from the OC community.

For the individual hypothesis we find that the hr-index had correlation numbers of (.200, -.401, .416) with Rp and (.442, .015, .416) with Pr (these are listed Degree, Closeness, and Betweenness Centralities). We also see that the hv-index had correlation numbers of (.394, -.529, .577) with Rp and (.649, .001, and .577) with Pr. Taking a look back at the hypotheses:

H1a: replies (hr) to degree centrality – Not supported for Rp, strong support for Pr.

H1b: replies (hr) to closeness centrality – Negative correlation for Rp. Not supported for Pr.

H1c: replies (hr) to betweenness centrality – Support for both Rp, Pr.

H2a: views (hv) to degree centrality – Some support for Rp, stronger support for Pr.

H2b: views (hv) to closeness centrality – Negative correlation for Rp. Not supported for Pr.

H2c: views (hv) to betweenness centrality – Strong support for both Rp, Pr.

6.8. Summary

This research continues the stream of research to find ways to measure user influence in an OC. Previous studies introduced the use of the Hirsch metric to measure the user contributions to the OC. We are trying to create a basket a measures to help identify different influence measures of a user of an OC. This study extended this measure by trying to include SNA measures to the mix of measures. We were able to confirm some support in some measures in this exploratory study.

We found that betweenness centrality was highly correlated with the hr and hv indices. This may be explained by the fact that there might be a cluster of high influence users in the middle of the social network. We also found negative correlation between Rp closeness centrality and the hr and hv-indices. While Pr closeness centrality was of no correlation. Finally we found stronger correlation between Pr degree centrality than Rp degree centrality. This is probably because of the reads going in the same direction as the measure of hv.

There are a few possible challenges in the use of the h-index in this study. First the use of the h-index can be problematic. The h-index for the OC is inherently higher than those of the bibliometric kind. The hv-index is measuring the display of the text to screen and not really a ‘read’. The content of replies are not read by the study so a reply may not even be relevant to the original topic. A reply may actually be a reply to a subsequent poster and not the original thread creator so the influence may be measured incorrectly. While the administrators do filter out off-topic postings so this weakness may be minimized. We also realize that the h-indices take time to garner a number so only veteran users are going to show up on the high h lists.

Table 6-1. The Correlation Matrix

		RpDegree	RpClose	RpBetwn	PrDegree	PrClose	PrBetwn	Hr	HV
RpDegree	Pearson Correlation	1	-.266**	.412**	.297**	-.162**	.412**	.299**	.394**
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.000
	N	1641	1641	1641	1640	1641	1641	1641	1641
RpClose	Pearson Correlation	-.266**	1	-.356**	-.389**	-.154**	-.356**	-.401**	-.529**
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.000	.000
	N	1641	1641	1641	1640	1641	1641	1641	1641
RpBetwn	Pearson Correlation	.412**	-.356**	1	.827**	-.039	1.000**	.416**	.577**
	Sig. (2-tailed)	.000	.000	.	.000	.113	.	.000	.000
	N	1641	1641	1641	1640	1641	1641	1641	1641
PrDegree	Pearson Correlation	.297**	-.389**	.827**	1	-.027	.827**	.442**	.649**
	Sig. (2-tailed)	.000	.000	.000	.	.283	.000	.000	.000
	N	1640	1640	1640	1640	1640	1640	1640	1640
PrClose	Pearson Correlation	-.162**	-.154**	-.039	-.027	1	-.039	.015	.001
	Sig. (2-tailed)	.000	.000	.113	.283	.	.113	.555	.979
	N	1641	1641	1641	1640	1641	1641	1641	1641
PrBetwn	Pearson Correlation	.412**	-.356**	1.000**	.827**	-.039	1	.416**	.577**
	Sig. (2-tailed)	.000	.000	.	.000	.113	.	.000	.000
	N	1641	1641	1641	1640	1641	1641	1641	1641
Hr	Pearson Correlation	.299**	-.401**	.416**	.442**	.015	.416**	1	.672**
	Sig. (2-tailed)	.000	.000	.000	.000	.555	.000	.	.000
	N	1641	1641	1641	1640	1641	1641	1641	1641
HV	Pearson Correlation	.394**	-.529**	.577**	.649**	.001	.577**	.672**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.979	.000	.000	.
	N	1641	1641	1641	1640	1641	1641	1641	1641

** . Correlation is significant at the 0.01 level (2-tailed).

The calculation of the hv-index and the social network of reads are algebraically related. This may account for high correlations between the SNA numbers and the hv-index. Another challenge is the use of a response network may not be suitable for analysis. The use of digital trace data collected from OC’s has generated some caution as to how this data might be used and analyzed using SNA techniques (Howison et al. 2011).

We also realize this a really specific type of OC, and therefore cannot generalize to other OC’s. We do believe the methodology of using both h-index and SNA measures can be a benefit to OC’s in general.

We also realize the data was taken over a short period of time and that more data over more seasons might show us better and more insightful results.

Some possible future research areas for this study include (1) the use of other SNA measures. (2) The addition of more seasons of data points. (3) Exploring if this methodology can be used on other similar type of OC's and other completely different types of OC's.

6.9. Conclusion

The current study is a part of a stream of research in examining how to measure the influence of user contribution in an OC. This is important for the stakeholders of the OC including the users, administrators, moderators, and owner of the OC. We added the social network measures to previously introduced measures of h-indices to measure the influence of OC users. We were able to show some correlation between these measures.

Chapitre 7

Measuring New Member Contribution in an Online Community: A Preferential Attachment Approach

Abstract

The online community (OC) is a popular form of specialized knowledge transfer, where geographically dispersed users can form a community to share ideas, send messages, and forge online friendships. OC's tend to have a life cycle, with the birth and growth of the OC followed by a stagnant stage where users stop posting and the OC eventually dies due to inactivity. One way that an OC can stay vibrant is to encourage new contributors. In this research in progress, we evaluate new users of an OC for preferential attachment, where new users tend to interact and post on thread starts of authors that are well known and respected in the online community. We look at the existing users' posting levels to measure veteran users' impact and measure new user behavior compared to veteran user of the OC. We operationalize our method using a college sports team fan site.

Keywords: contributor evaluation, new user evaluation, online communities, preferential attachment

7.1. Introduction

An online community is defined as an “aggregation of individuals or business partners who interact around a shared interest, where interaction is at least partially supported and/or mediated by technology and guided by some protocols or norms” (Porter 2004). An OC may also be referred to as a virtual community, an internet community, message board, or forum. Online communities (OC) continue to grow in popularity and sophistication as a means for geographically dispersed members who have similar interests to communicate, share opinions and multimedia. OC’s offer a way for people with common interests to actively participate in sharing of information, debate relevant (or even irrelevant topics), forge online connections, and build a sense of being part of a vibrant community. OC’s are often used as a source of knowledge and a repository of information from which members draw on their real-life experiences. For these communities to exist, active participation by the members is necessary. The contributions by members also need to be relevant, correct, and up-to-date.

Sometimes an OC is tied to an entity that in itself can become popular but over time diminish in popularity. There are many OC’s dedicated to the gaming or sports community where the product or franchise may have been popular at one time but the product may not have had updates or the team may have moved (e.g. Red Dead Redemption or the Atlanta Thrashers). There are also many products that have been kept updated and the OC’s tied to these products tend to live on (e.g. Halo, World of Warcraft).

OC’s tend to have a lifecycle with an initial acceleration in participation, a saturation point where participation seems to plateau, and a dying off period where the community loses interest in the OC (Iriberry and Leroy 2009) This dying off may be due various reasons such as user interest or the decrease in popularity of the topic. Fighting against this lifecycle tendency is the OC owner attempting to keep the OC vibrant and finding new ways of engaging community members. This may involve promoting more relevant discussions, highlighting important member contributions, and creating new content. Many OC’s are supported through advertising revenue. In these OC’s, the site is monetized through a mixture of views, click-through rates (CTR) on ads, and purchases through ads. One such site is the topic of our research paper, BigU – an online community for a university sports team that will be introduced in more detail later.

For our research stream, we have an overall objective of addressing OC profitability. To sustain an OC, an OC owner must cover the expenses of hosting, creating content, and maintaining the OC. Monetizing the site means, at the very least, many people must see the advertisements. However, to

attract the people to the OC, content also must be stimulating and relevant. Thus, our overall goal became a more targeted research question: How do OC owners uphold the vibrancy of an online community? For the purpose of this research, vibrancy refers to the interactivity in the OC. Vibrancy is a measure of the postings and subsequent views and replies to these postings in the OC. A vibrant OC, then, is an OC that has frequent posting submissions and much interaction with those posts. When an OC loses vibrancy the OC interaction become stagnant, postings become scarce; page views and reply activities on postings also cease. An OC needs to attract many people and that means user-created content must be interesting.

For an OC owner the addition of new members and encouraging those new members toward becoming major contributors is an important task. Yet, there are gaps in the current research around new user behavior in OC's. While online communities may have survived for a period of time, new ideas are necessary for the OC to grow and maintain vibrancy. While currently active users may continue to contribute, recruiting and gaining new contributing users is a good method to continue maintain OC vibrancy. For this study we focus on new contributors and the notion of preferential attachment (Barabasi, 2012; Faraj and Johnson 2011; Johnson et al. 2014), where new contributors tend to connect with historically high-end users, to explain new member contribution.

The rest of the article continues as follows. First, we explore the background of OC's, the notion of preferential attachment and how we measure influential users. Second, we present BigU and how the data collection process was conducted. Third we propose measures for new users contributions and see if they are consistent with preferential attachment. Finally, we present our preliminary findings and then finish with a discussion and conclusion.

7.2. Background

We investigate OC's and focus especially on OC health and vibrancy attainment in OC's. We then examine preferential attachment and how we derive our bibliometric measures.

7.2.1. Online Communities

OC's require interactions, contributions, and visits from the OC community. Unlike e-commerce websites selling tangible products, OC's requires major interaction from the customers (the community members) for revenue. The OC also can have a customer come back as a repeat customer over and over because the content changes rapidly. For example, in our case study OC, many users log in multiple times during the day because new content and news is posted throughout

the day. So long as the OC content continues to add value to the customer the OC owners can reap the benefits of providing this service.

There are many studies conducted on OC's (e.g. Harper et al. 2007; Iriberry and Leroy 2009; Ren and Kraut 2007; Sassenberg and Postmes 2002). OC's are usually formed around particular topics or areas of interest. Users join OC's for a wide variety of reasons including accessing more information or sharing information about their common interests. Research in OC's have targeted different types of OC's. For example, researchers have focused on movies and entertainment (Keen 1980; Wang and Fesenmaier 2003), travel (Wang and Fesenmaier 2004), business (Wang and Fesenmaier 2003; Wang et al. 2008), health (Wang and Fesenmaier 2003), and computers/science (Wang and Fesenmaier 2003; Zheng et al. 2007). In the IS field recent studies have examined OC's in the areas of knowledge management systems (Kankanhalli et al. 2005), open source software development (von Krogh et al. 2003; von Krogh et al. 2012), and packaged software development (Holström and Henfridsson 2006).

Current research on OC participation has evolved from the peer-to-peer (P2P) file sharing research. The P2P research tended to look at the contributor vs. downloader mix and how the community was riding on the work by the contributors (Golle et al. 2001). More recent research attempts to define some form of evaluation system to rate contributors and the impact of their contribution. For example, users that read the contributions might be solicited to provide feedback. The systems are typically set up to give credit to the contributor as well as the evaluator in the form of some sort of compensation (contribution points, virtual currency, etc.).

Whether an OC becomes and remains successful is heavily dependent on the participation by community members. The OC is only as good as the content created by the community members (Assmann et al. 2009; Cothrel et al. 1999). OC's tend to have a lifecycle where, in the beginning they are vibrant and relevant, a middle where they maintain vibrancy, and then an end where they fall out with the community and there are minimal posts and low interaction over a long period of time (Iriberry et al. 2009). OC's that die out tend to have irrelevant or outdated contributions and members find the content no longer interesting and stop visiting the OC. These events lessen new contributions and, thus, lead to a vicious cycle of stagnant behavior by the community members. Even if there are many visitors to the OC, visits alone do not guarantee vibrancy since community interaction is not static (Mousavidin et al. 2009). There are studies of OC's addressing the downfall of OC's (Gainer et al. 2003; Keen 1980, Koh et al. 2007; Rothaermel and Sugiyama 2001; Trevino 1999). OC's may appeal to visitors so long as new and interesting material is available. However, as communities decrease frequency of new contributions, there is less incentive for current members to visit the community on a regular basis. Once an OC loses momentum, trying to win back prior visitors

becomes more difficult. Many studies have tried to address this problem by trying to identify why the community becomes stagnant and identifying ways to improve the contributions to an OC (Gainer et al. 2003; Preece 2001; Ren et al. 2007; Rothaermel and Sugiyama 2001; Trevino 1999). Therefore, a need exists in practice to ensure the vibrancy of the OC is not lost. The challenge for OC owners is to try extending the initial growth or middle vibrancy period as long as possible to maintain the health of the OC. While many papers have been written about large OC's such as Myspace, Facebook, and ebay we have found that research looking into the health of smaller OC's is lacking (Clavio 2008; Shen et al. 2006).

There is also research concerned with the identification of users that contribute. The 'user status' is a measure of the community users used by many OC's that is derived from the number of postings by a user. Typically the 'user status' is a rank for a user that is given after the user has contributed a certain amount of posts or been an active member for a particular length of time. For example, in a sports related OC a user might gain 'bench player', 'rookie', 'regular', 'all-star' status by posting 5, 10, 100, and 1000 times, respectively. User status in OC's has been an area of research in the past. This area relies on social capital (Ganley et al. 2006; Law et al. 2008), social presence (Shen et al. 2009; Shen et al. 2006), social identification (Shen et al. 2008; Zheng et al. 2007), and social influence theory (Assmann et al. 2009; Cheng et al. 2006). Some of these studies try to make the connection between high status users and the vibrancy and thus the health and success of the OC (Assmann et al. 2009; Cheng et al. 2006).

Some studies have evaluated why there is a loss in community member activity and how to help stagnant communities become more active (Harper et al. 2007). Several studies have shown that lurkers, those that are members and benefit from the OC but do not contribute, exist in all OC's and that this form of 'social loafing' is detrimental to communities (Shiue et al. 2010). Some studies identify 'loafing' or 'lurking' as a contributor to the demise of an OC (Dantzig 2002; Gainer et al. 2003, Keen 1980; Rothaermel and Sugiyama 2001). Loafing is "the tendency for individuals to expend less effort when working collectively than when working individually" (Shiue et al. 2010 p. 769). Lurking is the activity of reading "messages on an Internet discussion ... without contributing" (Merriam-Webster Online Dictionary 2010). Member non-participation may indicate dissatisfaction and the lack of a compelling reason for the lurker to become active. Lurkers may be less satisfied than active posters or do not want to add confusion to already-busy message boards (Preece 2004).

There has been some research trying to change the 'lurkers' into 'contributors'. These studies have tried to identify the motivation for contributors to contribute. The reasons for contributors to contribute have been identified as being a way to build up one's reputation online (Kankanhalli et al.

2005; Ma and Agarwal 2014; Ren et al. 2012; Wasko and Faraj 2005) and/or contributors contribute knowledge just because it is satisfying to help others (Kankanhalli et al. 2005; Wasko and Faraj 2005). Other studies have looked specifically at the feedback mechanism to allow the contributors to know when they are gaining reputation in the OC (Burke et al. 2009; Cheshire and Antin 2008; Jung et al. 2010).

7.2.2. Preferential Attachment

Preferential attachment first mathematically appeared in 1923 with the urn model by Eggenberger and Polya (as cited by Barabasi and Albert 1999). The mathematical model had been repeatedly appeared but the model was termed the “Matthew Effect” by Robert Merton in 1968. The “Matthew Effect” is where “the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark” (Merton 1968). Merton observed that co-authors of scientific papers did not have equal reputation and that the author with the higher reputation got more recognition than the less reputable co-authors.

The preferential attachment notion has translated into network analysis. “A node joining a network, such as a new web page or a new protein, can in principle connect to any pre-existing node. However, preferential attachment dictates that its choice will not be entirely random, but linearly biased by the degree of the pre-existing nodes – that is, the number of links that the nodes have with other nodes” (Barabasi 2012, p. 507). This means that in an established network “new vertices attach preferentially to sites that are already well connected” (Barabasi and Albert 1999 p. 509). The implication of preferential attachment for online communities is that new entrants into the online community will attach themselves to established community members. Preferential attachment “occurs in a network when new actors choose to interact with already well-connected others over more typical others” (Faraj and Johnson 2011 p. 1468).

Users in an OC typically have access to status about other users in the OC. In “many online communities, participants can readily assess the relative characteristics of existing participants” (Faraj and Johnson 2011 p. 1469). For “a social setting such as online interaction, the implication is that individuals prefer to initiate communication with those members that are highly active and visibly well connected” (Ibid p. 1469). Faraj and Johnson (Ibid) used preferential attachment by observing five online communities and running simulation to explain the contributions of new users. For their proposition “Online communities demonstrate a structural network tendency toward preferential attachment” (Ibid p. 1469) they found “evidence that this tendency exists in the

networks in the opposite direction as proposed” (Ibid p. 1473). They found contrary to previous models on social networks, “not only a lack of evidence for preferential attachment, but rather a disinclination toward preferential attachment” (Ibid p. 1475). Given this surprising finding from Faraj and Johnson (Ibid) we propose to test their findings using data from an OC.

7.3. Propositions

For our OC we have new users entering the OC all the time. All users also have the ability to see the status of other users. Although the OC uses ‘user status’ – a membership level status which is automatically updated daily and is solely dependent on posting counts for “normal” members or an indication of moderator level for staff members – the users can also see the up-to-date actual number of posts made by other users. Given this, new users may lurk to identify who are the reputable users on the OC. Anyone on the OC can post to the OC, whether it be replying to an active thread or starting their own new thread. For a new user that wants to connect to a reputable user, the choice should be to reply to a post by a reputable user. If the new user does not want to use preferential attachment, they can test the waters by posting a new thread start on their own. For our propositions we define new users as users that have not posted in the past while veteran users as users that have posted in the past. Given the findings of Faraj and Johnson (2011) we propose the following:

P1: New people would have a tendency to start new threads rather than attach to established people of repute within the network compared to veteran users.

P2: As people become socialized into the community, veteran users will tend to reply to more threads than new users.

P3: Veteran users would reply to users with a higher number of posts compared to new users.

7.4. Research Context

This study focuses on the case of BigUFans , an independently owned and operated website that covers and provides an OC of like-minded fans of a large university in the United States. The site owner (one of the authors is an alumnus of Big U. and the owner of BigUFans.com) has been attempting to better understand the behavior of new users on the OC and to more accurately

identify those who might be receptive to becoming high-level contributors to the OC. The owner realizes that maintaining a core group of users who participate regularly is the key to success in an online community. Other members typically wait until these core members post before they themselves reply. BigUFans is representative of many similar sites, which court advertisers to a hard-core, niche group of dedicated followers. It is through advertising that the site has, over 15 years, earned enough revenue to keep the site online and improve its server capabilities, which, in turn, have made the site more usable. However, it is the vibrancy of the community that makes it relevant.

Big U. is a large tier I research institution in the southeastern US. Big U. is a public state university that has been around for over 150 years and now has over 30,000 students enrolled. The OC is focused on generating content and user contributions about the athletic teams of Big U. where football is a major sport. Big U competes in the NCAA's Football Bowl Subdivision (formerly Division I) in many athletic disciplines including the major sports of football, basketball, and baseball.

Currently BigUFans.com has four major competitors. BigUFans.com was created in 1999 and is the largest completely free OC out of the five competing OC's. The OC generates revenue through advertising sales based on page views and advertisement click-through rates. Thus, sustained user activity on BigUFans.com is critical for revenue generation. BigUFans.com's competitor sites have a combination of paid access to content and advertising. Conversely, anyone may anonymously browse all content on BigUFans.com but must be a member (registration is free) to post a reply or start a new thread in the BigUFans.com forums.

There are six administrators and 15 moderators of BigUFans.com that keep the topics on the forum of BigUFans.com relevant and free of spam and incendiary comments. Registered members can become moderators as they become more active in the community but only at the invitation of the owner of the site. Because of the work of the staff, one of the major advantages of BigUFans.com over its competitors is that the community feels the OC has more positive messages that are free of advertising, spam, flame wars, and off-topic postings.

The majority of activity on BigUFans.com comes from discussion about the football program. The site uses Google Analytics to gain insight into visitor demographics and site activity. During the 2014-2015 season monthly activity saw an average of 1.5 million page views, 801 new posts, 167,000 unique visitors, and 85 topics started. Daily activity included 600 signed in users (about 5% of all visitors) and 2 new users. A typical visit consisted of a viewing of 5.1 pages and a duration over 6 minutes. The demographics were largely male (54%), college graduate (72%), and of middle to upper income (71%) and between 18 and 45 (75%). During the offseason prior to the 2014-2015 season, there were 700,000 page views a month.

Currently BigUFans.com gives different level user status automatically to users that post a certain amount, regardless of how many users read their messages. Currently BigUFans gives the following user status (along with minimum posts required); Scout Team (1 minimum post required), Third team (13), Second Team (50), First Team (250), All-Conference (1000), All-American (2000), and Hall-of-Fame (5000). Currently there are 20470 Scout Team members, 1363 Third Team members, 1086 Second Team members, 590 First Team members, 179 All-Conference members, 121 All-American members, and 47 Hall-of-Fame members. The Scout Team is basically given to anyone that signs up for the service. Both 'user status' and 'number of posts' of all users are visible to all others users of the OC.

7.5. Research Methodology

We collected data from the site message board database to include all message board posts in the calendar year 2014. New users were defined as users that signed up in 2014. Veterans were considered those users that signed up before January 1, 2014. This included users from the inception of the OC in April 1999 through December 2013. Posts include both replies to existing threads and new thread starts, and these are mutually exclusive. The data yielded 213,050 unique posts, of which 5,297 were from new users. Within the posts there were 6,136 unique thread starts, of which 191 were by new users. There were 1,441 unique users who posted in 2014, of which 103 were by new users; 1,337 unique veteran users posted, of which 609 created new threads; 103 unique new users posted, of which 61 created new threads.

Data for each of the posts included a unique post ID and Post Creator ID, the post creation Date, join date of the poster, and number of all-time posts for the user. For each post we were able to distinguish between a reply to a thread or thread start, the unique poster, and to whom they replied along with the original author's number of all-time posts. This allowed us to distinguish between new users and veterans and see whether they started threads, or replied to threads.

7.6. Results

The following tables show the results of our preliminary work on this study. Table 7-1 shows the counts of posts (replies to existing threads) and thread starts for veterans (users signed up before January 1, 2014) and for new users (users signed up in 2014). The ratio shows that the new users tended to create more thread starts compared to veteran posters. Table 7-1 results show a support for propositions 1 and 2.

Table 7-1. Count of Posts and Ratio

	Veterans	New Users
Posts (reply to thread)	201619	5298
Thread Starts	5945	192
Ratio (Thread Start/Post)	2.86%	3.49%

We found that some new users became prolific posters in 2014. So perhaps we would see preferential attachment effects on early postings of new users. In particular we wanted to look at the first posting of a new user. Table 7-2 shows just the first posting of the new users that posted in 2014. Of the 103 new users that posted in 2014, we see that 31% of those were thread starts. Surprisingly, new users seemed much more confident in starting a thread right off the bat, even more so over the long run (compared to the 3.49% in table 7-1). We found many new users actually only posted one time. So table 7-2 also shows for when we filtered out those new users that posted once and disappeared. We found 28 new users from 2014 that posted only one time in 2014, and it was evenly split between replies and thread starts.

Table 7-2. First posts by new user and ratio

	New Users	Filtered for one time posters
Posts (reply to thread)	71	32
Thread Starts	32	18
Ratio (Thread Starts/Posts)	31%	24%

Finally we wanted to see the strength of preferential attachment. For all the replies to threads that existed, preferential attachment would assume that new users would tend to find threads started by veterans with a large number of posts in the past. Table 7-3 shows the average number of posts of the thread starter for all posts (replies) in 2014, separated by the status, veteran or new user, of the replier. We found that veterans would reply to threads started by users with a higher number of posts in the past compared to new users. Table 7-3 shows initial support for proposition 3 that veteran users would reply to users with a higher number of posts compared to new users.

Table 7-3. Average of thread starter number of posts for replies

	New Users	Veterans
Average of thread starter number of posts	15631	21364

7.7. Discussion and Implications

When we first began this research we believed that with preferential attachment new users would attach themselves to veteran users. We were surprised by the work by Faraj and Johnson (2011), which had findings that contradicted preferential attachment. Indeed the findings “undermine previous explanation that online exchange follows a power law distribution based on people wanting to connect to ‘popular’ others in online communities” (Ibid p. 1464). Our findings confirm the findings of Farah and Johnson. We find that there is a ‘disinclination’ by the new users to go out of their way to not attach to reputable users. One explanation is that in OC’s as “social communication networks differ from physical and technological networks studied by many network researchers” (Ibid p. 1475). Our study is also taking data from one highly specialized OC, which does not necessarily imply generalizability to other OC’s.

Some limitations to our study include the sample size, the difference in development of the new user over a year period, and generalizability of the topic. The current study found only 103 new users who posted in 2014. This is dwarfed by the size of the total of 474 user registrations and the 1441 unique users who posted, overall. We also defined new users to be those users that signed up and posted in 2014. While some users may have been new in December of 2014, some were also new in January 2014. These two ‘new’ users sign up dates were separated by almost a year yet they were both counted as ‘new users’. Some users that signed up early in 2014 went on to become prolific contributors to the OC, so much so that their behavior should not have counted as ‘new users’ to the end of 2014. Finally the OC we used was highly specialized but also had an advantage of being a sports team, which meant that there were many topics of discussion over the course of a season. For example, an OC dedicated to a particular computer game may not have additional topics after a player has concluded the game.

In light of these limitations our future studies may include exploring other types of OC’s, taking a stratified view of the OC, and using other sources to measure user status. Future studies will look into various types of OC’s to examine the impact that the type of OC has on new member contributions. Another study may be to change the methodology to look at historical data from the past and see if there are any stratified differences in the age of the user. For example, new users may be measured by not calendar year as we did but time elapsed since sign up. There might be differences in new user behavior between the first week, first month, first three months, first year, and so forth. We may see behavioral differences between veterans that have been on the OC for two years versus five years. Finally we may use measures other than posting counts to measure the reputation of a veteran user, such as using a hybrid h-index.

7.8. Conclusion

We find that the current study can have implications for practitioners to understand the behavior of new users to an OC. Different types of new users may be revealed in our future studies. Using this information that new users tend to start new threads, practitioners may want to encourage new users to join and in essence create 'new blood' to expand on the OC.

We commenced this research to find new methods for OC owners to find users that contribute to the vibrancy of the OC. We modeled our preliminary study after the Faraj and Johnson (2011) study and applied preferential attachment on a university sports team OC and our findings were consistent with them in that new user contribution seemed to have a inclination away from preferential attachment. New users of our OC are not attaching to reputable users of the OC. This finding raises the same questions that Faraj and Johnson (Ibid) brought up about past research tied to power distribution and preferential attachment and calls for additional studies in this area using more relevant measures of reputation and impact.

Chapitre 8

Measuring Member Contribution Impact in an Online Community: A New Hirsch Index Inspired Approach

Abstract

The online community is a popular form of specialized knowledge transfer, where geographically dispersed users can form a community by sharing ideas, send and post messages, debate topics, and forge online friendships. One of the problems with online communities is that they tend to have a life cycle, where there is the birth and growth of the online community but then there is a stagnant stage where users stop posting to the online community and the community eventually dies due to inactivity. Trying to extend the vibrant growth stage of an online community is a relevant topic for any administrator of an online community. One way that an online community can stay vibrant is to encourage contributions.

In this research, we propose using a non-invasive bibliometric measure using the Hirsch-index methodology as a way to identify high-level contributors. We proposed that those users that are high in perceived identity verification in the form of replies and views to threads started would have high contribution in the form of the Hirsch index measure. We ran the data on a college sports fan site. We also analyzed the data to identify high-level contributors to the online community. Finally we demonstrate how this new measure can be used using a real online community.

Keywords: h-index, online communities, contributor evaluation

8.1. Introduction

Online communities (OC) continue to grow in popularity and sophistication as a means for geographically dispersed members that have similar interests to communicate and share ideas and thoughts and multimedia. These OC's offer a way for people with common interests to actively participate in sharing of information, debate relevant (or even irrelevant topics), forge online connections, and build a sense of being part of a vibrant OC. OC's are often used as a source of knowledge and a repository of information from which members draw on their real-life experiences. But in order for these communities to exist, there needs to be active participation by the members. The contributions by members need to also be relevant, correct, and up-to-date.

An online community is defined as an "aggregation of individuals or business partners who interact around a shared interest, where interaction is at least partially supported and/or mediated by technology and guided by some protocols or norms" (Porter 2004). Sometimes an OC is referred to as a virtual community or an Internet community. These OC's can be gathering places for people with similar interests where they can share information, talk about experiences, or just meet up with friends to chat.

While many OC's pop up all the time, we have seen many OC's that have died off in the past. Many of the no longer relevant OC's can be seen as a specialty community in Yahoo Groups or Google Groups. You may have seen these as they tend to have forum questions that have been closed due to inactivity or discussions that were lively at some point but tend to have little conversations in the past few years. Other now less popular OC's such as Second Life and Myspace are more categorized as a social networking site but have OC community characteristics and can have OC components to them. These social networking sites have the ability to post information and share information between users which is a common characteristic of an OC. Vibrant and thriving OC's such as Tripadvisor, Flyertalk, and Rotten Tomatoes also exist. In each of these sites there are ways to report reviews of products of interest, i.e. hotels for Tripadvisor or movies for Rotten Tomatoes, and users can give feedback to the reviews, whether it was a helpful review or not.

Sometimes an OC is tied to an entity that in itself can become popular but over time diminish in popularity. There are many OC's dedicated to the gaming or sports community where the product or franchise may have been popular at one time but the product may not have had updates or the team may have moved (e.g. Red Dead Redemption or the Atlanta Thrashers). There are also many products that have been kept updated and the OC's tied to these products tend to live on (e.g. Halo, World of Warcraft).

While there may be a life-cycle effect for an OC, an OC owner's goal is to keep the OC vibrant and to engage community members. This involves having topics and discussions that are relevant and timely. Many OC's need to gain revenue via advertising and the advertising dollars come in because of a mixture of views, click-throughs on ads, and purchases through ads. One such site is the topic of our research paper, BigU, which is an online community for a university sports team that will be introduced in more detail later.

For our research stream we have an overall objective of finding ways for keeping an OC up and running. To do this the OC must maintain revenue to cover the expenses of hosting, creating content, and maintaining the OC. To keep revenue up the OC must maintain a good advertising base, which means, at the very least, many people must see the advertisements. However, to attract the people to the OC content must be stimulating and timely. So our overall goal became a more targeted research question: How does one keep the vibrancy of an online community? For the purpose of this research, vibrancy refers to the interactivity in the OC. Vibrancy is a measure of the postings and subsequent views and replies to these postings in the OC. A vibrant OC is then an OC that has frequent posting submissions and much interaction with those posts. When an OC loses vibrancy the OC interaction become stagnant, postings become scarce, and view and reply activities on postings also cease. To keep the vibrancy an OC needs to attract many people and that means that user-created content must be interesting.

One problem area that we found is that there is a lack of ways to identify users that contribute relevant content for the community. There is the simple 'posting number' that many OC's use today to demonstrate how much a user has contributed to an OC. Unfortunately, post counts are simply a measure of productivity. Impact, influence, and whether other users are reading their contributions has no bearing in a pure count of posts. Therefore, the current research is attempting to find better ways to measure the influence that contributors have on the online community. We propose an aggregate measure of using social network analysis and bibliometrics to measure the influence and productivity of a user, and run data from an online community. In this exploratory research we find different types of users that were identified by this aggregate measure as opposed to only looking at the 'posting number' measure.

The rest of the article continues as follows. First, we explore the background of OC's and some techniques in measuring influence from other disciplines. Second, we present BigU and how the data collection process was conducted. Third we propose measures for contributor influence. Finally, we present our findings and then finish with a discussion and conclusion.

8.2. Background

We first take a look at OC's focusing especially on OC health and how vibrancy of an OC is attained. We then look at two main areas where we derive our measures, bibliometrics and social network analysis.

8.2.1 Online Communities

Our research is concerned with the health of an OC. We are also interested in how to maintain this health and vibrancy of the OC. OC's require interactions, contributions, and visits from the OC community. Unlike a e-commerce website selling a tangible product, an OC requires major interaction from the customers (whom are the community members) for its revenue stream, but the OC can also have a customer come back as a repeat customer over and over because the content can change so quickly. For example, in our case study OC, many users log in multiple times during the day because new content and news is posted throughout the day. So as long as the OC content continues to add value to the customer the OC owners can reap the benefits of providing this service.

There are many studies conducted on OC's (Harper et al. 2007; Iriberry and Leroy 2009; Ren and Kraut 2007; Sassenberg and Postmes 2002). OC's are usually formed by some topic/interest area. The users join the OC in order to get more information or share information about their common interest. Research in OC's have targeted various different types of OC's. For example there is research that has targeted sites on topics such as movies/entertainment (Keen 1980; Wang and Fesenmaier 2003), travel (Wang and Fesenmaier 2004), business (Wang and Fesenmaier 2003; Wang et al. 2008), health (Wang and Fesenmaier 2003), and computers/science (Wang and Fesenmaier 2003; Zhang et al. 2007). In the IS field there are studies that looked at knowledge management systems (Kankanhalli et al. 2005), as well as studies that looked at software development, addressing OC's that are concerned with open source software development (von Krogh et al. 2003; von Krogh et al. 2012), as well as packaged software development (Holström and Henfridsson 2006).

The current research on OC participation has evolved from the peer-to-peer (P2P) file sharing research. The P2P research tended to look at the contributor vs. downloader mix and how the community was riding on the work by the contributors (Golle et al. 2001). More recent research attempts to define some form of evaluation system to rate contributors and the impact of their contribution. For example, users that read the contributions might be solicited to provide feedback. The systems are typically set up to give credit to the contributor as well as the evaluator in the form of some sort of compensation (contribution points, virtual currency, etc.).

Whether an OC becomes and remains successful is heavily dependent on the participation by community members. The OC is only as good as the content created by the community members (Assmann et al. 2009; Cothrel et al. 1999). OC's tend to have a lifecycle where, in the beginning they are vibrant and relevant, a middle where they maintain vibrancy, and then an end where they fall out with the community and there are minimal posts and low interaction over a long period of time (Iriberry et al. 2009). OC's that die out tend to have irrelevant or outdated contributions and members find the content no longer interesting and stop visiting the OC. These events lessen new contributions and, thus, lead to a vicious cycle of stagnant behavior by the community members. Even if there are many visitors to the OC, visits alone do not guarantee vibrancy since community interaction is not static (Mousavidin et al. 2009). There are studies of OC's that addressing the downfall of OC's (Gainer et al. 2003; Keen 1980; Koh et al. 2007; Rothaermel and Sugiyama 2001; Trevino 1999). OC's may appeal to visitors so long as new and interesting material is available. However, as communities decrease frequency of new contributions, there is less incentive for current members to visit the community on a regular basis. Once an OC loses momentum trying to win back prior visitors becomes much more difficult. Many studies have tried to address this problem by trying to identify why the community becomes stagnant and identifying ways to improve the contributions to an OC. (Gainer et al. 2003; Preece 2001; Ren et al. 2007; Rothaermel and Sugiyama 2001; Trevino 1999). Therefore, a need exists in practice to ensure the vibrancy of the OC is not lost. The challenge for OC owners is to try extending the initial growth or middle vibrancy period as long as possible to maintain the health of the OC. While many papers have been written about large OC's such as Myspace, Facebook, and ebay we have found that research looking into the health of smaller OC's is lacking (Clavio 2008; Shen et al. 2006).

There is also research concerned with the identification of users that contribute. The 'user status' is a measure of the community users used by many OC's that is derived from the number of postings by a user. Typically the 'user status' is a rank for a user that is given after the user has contributed a certain amount of posts or been an active member for a particular length of time. For example, in a sports related OC a user might gain 'bench player', 'rookie', 'regular', 'all-star' status by posting 5, 10, 100, and 1000 times, respectively. User status in OC's has been an area of research in the past. This area relies on social capital (Ganley et al. 2006; Law et al. 2008), social presence (Shen et al. 2009; Shen et al. 2006), social identification (Shen et al. 2008; Zheng et al. 2007), and social influence theory (Assmann et al. 2009; Cheng et al. 2006). Some of these studies try to make the connection between high status users and the vibrancy and thus the health and success of the OC (Assmann et al. 2009; Cheng et al. 2006).

We argue that 'user status' is only a count of production and by no means is it a measure of how the postings are received by the community members. One can post thousands of posts but those posts may be read by no one in the OC, thus not really contributing anything to the health and vibrancy of the OC.

Some studies have evaluated why there is a loss in community member activity and how to help stagnant communities become more active (Harper et al. 2007). Several studies have shown that lurkers exist in all OC's and that this form of 'social loafing' is detrimental to communities (Shiue et al. 2010). Some studies identify 'loafing' or 'lurking' as a contributor to the demise of an OC (Dantzig 2002; Gainer et al. 2003; Keen 1980; Rothaermel and Sugiyama 2001). Loafing is "the tendency for individuals to expend less effort when working collectively than when working individually" (Shiue et al. 2010 p. 769). Lurking is the activity of reading "messages on an Internet discussion ... without contributing" (Merriam-Webster Online Dictionary 2010). Member non-participation may indicate dissatisfaction and the lack of a compelling reason for the lurker to become active. Lurkers may be less satisfied than active posters or do not want to add confusion to already-busy message boards (Preece 2004).

There has been some research trying to change the 'lurkers' into 'contributors'. These studies have tried to identify the motivation for contributors to contribute. The reasons for contributors to contribute have been identified as being a way to build up one's reputation online (Kankanhalli et al. 2005; Ma and Agarwal 2007; Ren et al. 2012; Wasko and Faraj 2005) and/or contributors contribute knowledge just because it is satisfying to help others (Kankanhalli et al. 2005; Wasko and Faraj 2005). Other studies have looked specifically at the feedback mechanism to allow the contributors to know when they are gaining reputation in the OC (Burke et al. 2009; Cheshire and Antin 2008; Jung et al. 2010).

8.2.2 Invasive Methods

Two types of methods frequently occur in the literature: survey studies and evaluation studies. Both types require some form of user input. Feedback studies have some inherent feedback mechanism built into the OC. Typically these systems include functionality allowing users to 'rate' the work of another user or the target unit of measure of the website. For example, some OC's have the ability for a user to rate a contribution by another user. So if there was a question posed by a user and several people answered the question to varying degrees of detail, a user would be able to rate each of the answers given. While it is possible to have a multi-level rating system, typically these types of sites would only allow a 'Yes/No' type of Boolean rating system (Gainer et al. 2003; Keen 1980;

Trevino 1999). An example of a study conducted on a unit of measure was one study that targeted a movie review website (Keen 1980). Feedback was used by other members of the website to see a composite rating and user comments. These feedback systems typically allowed users to see others ratings to allow users to gain information to make their own decisions about different contributions (Gainer et al. 2003; Rashid et al. 2006; Trevino 1999).

Survey studies ask users questions to help users understand the importance of the community and try to understand why they were not contributing to the OC (Dantzig 2002; Koh et al. 2007; Rothaermel and Sugiyama 2001; Wang and Fesenmaier 2003). Some of these studies required some form of incentive to 'rate' another's work, or would solicit 'ratings' from a target audience. These included emails that were sent to the users imploring them to contribute to the OC (Keen 1980). Methodologies in these studies were varied but centered mostly on surveys. Some other methods seen were field experiments (Keen 1980) and experiments (Rashid et al. 2006). Other studies explained OC's by giving an overview (Skretta 2007), history (Skretta 2007), anthropology (Wilson and Peterson 2002), time series explanation (Wang et al. 2008), and success measures (Preece 2001)

8.2.3 Bibliometrics

This paper extends our research of using the Hirsch index to measure the user status and posting productivity (citation masked for review). The Hirsch index has been a measure used in bibliometrics that measures both productivity and influence of the authors. The old measure of author influence has been using survey research (Hirsch 2005). In the OC research area we see survey research targeting the owners or community members of small OC's (Law et al. 2008; Marett et al. 2009, Schaedel et al. 2010; Shen et al. 2008; Shen et al. 2009; Zheng et al. 2007). Survey research which rely on the ability of the survey taker to assess the OC contributions has been problematic in several ways. First, the survey taker has opinions and feelings that might bias the survey results. Second, the survey instrument is inexact. Typically a survey relies on measures like 5 point or 7 point Likert scales. This inexact measure introduces measurement error on the part of the survey taker (i.e. whether to put 5 or 6 for a question) and by the researcher interpreting the data. Third, the survey respondents may be a self-selected group that biases the data. Those people that answer surveys may be representative of a group of people that correlate with people likely to respond to a survey in the dataset. On the other hand the non-responders of surveys may be a group that is not represented in the dataset. For example lurkers tend not to post or reply to posts. Lurkers would likely not answer a request to fill out a survey. Thus lurkers may be a group that is not represented in a survey study. We propose using a bibliometric measure that looks at influence by quantitatively looking at actual replies and views (i.e. reads or displays to the computer screen).

The Hirsch-index was proposed by Hirsch in 2005 and is defined as “A scientist has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each.” in order to find the h -index of an author you would gather all the publications by said author and the citations to those publications. Then you would list the publications in descending citations order. When the i -th publication from the top has less than i citations you have found the h -index for that author.

The Hirsch index has been proposed as a way to look at OC postings in the past (Gomez et al. 2008; Laniado et al. 2011). In Gomez et al. (2008), they look at the discussion threads on Slashdot, a popular website with technology news. The researchers take the postings and developed a “measure to evaluate the degree of controversy provoked by a post” (Gomez et al., 2008). They do this by taking each initial post and create a tree from the initial post. The branches of the tree are responses to the initial post and the subsequent responses to the responses. The Gomez et al. h -index measures the maximum nesting level i of the tree where i “has at least $h > i$ comments, or in other words, $h+1$ is the first nesting level i which has less than i comments” (Gomez et al. 2008).

The Laniado et al. (2011) paper looks at wikipedia discussion pages. They try to characterize the discussions on the comments section of wikipedia. Laniado et al. (2011) do this by looking at the depth of the discussion on an initial comment as well as the breadth of the discussion using a h -index like measure. Laniado et al. (2011) construct trees out of the discussion, similar to Gomez et al. (2008). Laniado et al. (2011) create the tree using the response comments, which become child nodes for the initial discussion entry. For Laniado et al. (2011) the “ h -index of the tree is then the maximum level, for which the corresponding number of nodes is greater or equal to the level number” (Laniado et al. 2011). This is similar to the Gomez et al. (2008) version of the h -index. One key difference between these authors definition of the h -index from the original h -index is that the unit of measure for these discussion pages is the initial entry, and not the authors of these entries. They are able to find controversial and impactful initial posts but they cannot identify the individual contributors that are responsible for these posts using their h -index.

We use the h -index measure to define the influence for a contributor to the OC. We interpret citations in the OC to be replies to posts made in the OC. Our use of the h -index is in line with the initial definition of the h -index in that we are trying to find the authors of impactful posts. We are not interested in finding the impactful postings but rather want to find the authors that are creating interesting content that is being read and commented on by a large number of other users of the OC. We are different from the use of the h -index of Gomez et al. (2008) and Laniado et al. (2011) in that those authors have used it to measure the posts controversy, while we kept the original Hirsch

(2005) h-index unit of measure of the author. The difference that we have between the Hirsch (2005) definition is that we use comments instead of citations to measure the impact of the author of the posting in an OC.

We also have the distinct advantage over bibliometric studies in that we can find out who has viewed (or at least displayed onto their computer) posts from other users in the OC. So in addition to the replies (citation) we have a second h-index measure that looks at views. We define these as being the hr-index (replies) and hv-index (views). So the hr-index is defined as a contributor has index hr if hr of his/her N_p thread starts have at least hr comments each, and the other ($N_p - hr$) thread starts have no more than hr comments each. We also define the hv-index similarly as a contributor has index hv if hv of his/her N_p thread starts have at least hv comments each, and the other ($N_p - hv$) thread starts have no more than hv comments each.

8.2.4 Social Network Analysis

We also utilized several measures from social network analysis (SNA). We used the degree, closeness, and betweenness centrality measures. Centrality in SNA refers to the notion that in a network there are nodes that are central to the network. That is in a social network there are individuals that are central to the group. If a person is central to the group that means there are more people connected to that person in the group. Degree centrality is this simple count of how many connections you have in the group (Freeman 1977). The theoretical fully connected person will have $n-1$ connections (where n is the number of people in the group) or connections to everyone in the group except for herself. When someone is fully connected the distance between them and everyone else in the group is one. (Acedo et al. 2006; Eaton et al. 1999; Henry et al. 2007; Otto et al. 2002; Polites et al. 2008; Vidgen et al. 2007). Degree centrality is a measure of how many people you know in the network.

Closeness centrality is a measure of how distant you are from other individuals in the group. The distance is measured by the minimum number of connections to reach that other person. So for one individual, you would first find the distance between them and all $n-1$ other individuals in the group. The closeness centrality then finds the mean of these distances (Polites et al. 2008; Vidgen et al. 2007). The closeness centrality is a measure of how close you are to the rest of the network.

Betweenness centrality is a measure of how much you are a connector between two people on a network. If you are part of a minimum distance path between two people you are the “friend that knows a friend”. So people need you to connect to whomever they want to connect to. You are one of the crucial intermediaries to allow connections between two people. The betweenness finds all the

shortest distances that go through a person and then sums these distances and divides by $(n-2)(n-1)$ (Vidgen et al. 2007). The betweenness centrality is a measure of how important of an intermediary you are to connect individuals in the network.

8.3. Research Context

This study focuses on the case of BigUFans, an independently owned and operated website that covers and provides an online community of like-minded fans of a large University in the United States. The site owner has been looking for better measures of user value besides the simple count of posts. The owner realizes that maintaining a core group of users who participate regularly is the key to success in an online community. Other members typically wait until these core members post before they themselves reply. However, the owner also realized that it was not just the quantity of posts that identified the core user group. Instead, in OC's like BigUFans, the quality of the post and the reputation of the member are factors in how engaged other members become on a particular topic. BigUFans is representative of many similar sites, which court advertisers to a hard-core, niche group of dedicated followers. It is through advertising that the site has, over nearly 15 years, earned enough revenue to keep the site online and improve its server capabilities, which, in turn, have made the site more usable. However, it is the vibrancy of the community that makes it relevant. A better understanding of which users are key to this vibrancy is the goal of this research.

BigUFans.com is a OC website for the school Big University (one of the authors is an alumni of Big U. and the owner of BigUFans.com). Big U. is a large tier I research institution in the southeastern US. Big U. is a public state university that has been around for 150 years with over 30,000 students. The OC is in particular geared towards the athletic teams of Big U. where football is the major sport. Big U competes in the NCAA Division I in many athletic disciplines including the major sports of football, basketball, and baseball.

Currently BigUFans.com has four major competitors. BigUFans.com was created in 1999 and is the largest free OC out of five competing OC's. The OC makes revenue by advertising sales. Thus gaining activity on BigUFans.com is an important revenue generating activity. The four other competitor sites have some aspect of paid areas in their OC. Anyone may anonymously browse the contents of BigUFans.com but in order to post or reply to forums a user must register.

There are six administrators and 17 moderators of BigUFans.com that work to keep the topics on the forum of BigUFans.com relevant and free of spam and incendiary comments. Registered members can become moderators as they become more active in the community but only at the invitations of

the owner of the site. Because of the work of the staff, one of the major advantages of BigUFans.com over their competitors is that the community feels the OC has more positive messages that are free of advertising, spam, flame wars, and off-topic postings.

The majority of activity on BigUFans.com comes from discussion about the football program. During the 2011-2012 season monthly activity saw an average of 2.9 million page views, 1075 new posts, 167,000 unique visitors, and 89 topics started. Daily activity included 1,017 signed in users and 5 new users. A typical visit consisted of a viewing of 5.2 pages. The demographics were typically male (92%), college graduate (72%), and of middle to upper income (71%). During the offseason prior to the 2011-2012 season, there were 2 million page views a month. Currently BigUFans.com gives different level user status automatically to users that post a certain amount, regardless of how many users read their messages. Currently BigUFans gives the following user status (along with minimum posts required); Scout Team (0 minimum posts required), Third team (13), Second Team (50), First Team (250), All-Conference (1000), All-American (2000), and Hall-of-Fame (5000). Currently there are 20470 Scout Team members, 1363 Third Team members, 1086 Second Team members, 590 First Team members, 179 All-Conference members, 121 All-American members, and 47 Hall-of-Fame members. The Scout Team is basically given to anyone that signs up for the service. In addition there are six Administrators and 17 Moderators. These are given to users that have established themselves as responsible members (i.e. no spamming of message boards) and are asked to join these higher ranks.

8.4. Research Methodology

We collected data from the site message board database to include all message board topics created between October 23, 2012 and April 21, 2013. This included over 14,200 unique threads by 801 distinct authors. This resulted in 437,094 reads from 2,203 logged-in members. It should be noted that only about ten to fifteen percent of users log in. Additionally during this period, the site received 2,370,864 visits and 12,617,934 page views and had an average visit duration of over seven minutes per visit.

Data for each of the threads included the unique threadID and unique ThreadCreatorID, the creation Date, total views of the thread, and number of replies. For each user, we collected the unique userID, the user status (which reflects the status achieved based on the user's number of posts), the user join date, last visit date, last activity date, and last post date. Data collected for threads read data included the ReaderUserID, ThreadID, read date, ThreadCreatorUserID, thread creation date, number of thread views, and number of thread replies.

The authors wrote a Java program to calculate h-indices and the UCINET software program was employed to calculate central tendency measures. SPSS was used to import the data and calculate the correlation values for analysis.

Once the data was collected two sets of data were required for the new set of measures. First for the h-indices all postings were listed via postID for the data collection timeframe. This included the creator user ID, the number of people that viewed the post, and the number of people that replied to the post. For the SNA all postings and all readers were listed for postings during the data collection timeframe.

8.5. Measures

Our past research has shown that both the h-indices and the centrality measures can be used for measuring the influence of one member of an OC. In this research we plan to introduce a single aggregate measure to identify those that are important in the OC using a real dataset. The first page of the data output is shown in table 8-1 in the appendix. In table 8-1, the old method of counting posts, all proposed eight measures with their ranking for the individual with the ID number are shown. At the far right is the 'Rank Sum' which is the aggregate of the sum of the rank numbers for the proposed eight measures. The 'Rank Sum' aggregate measure is described in detail below.

8.5.1 h-index measures

The h-index measures were made for two different types. The views and the replies. These were labeled at hr and hv and are shown in table 8-1. We took the data in excel format and ran it through a Java program developed by one of the authors. This then created an output file with user ID's and their hr-index and hv-index.

8.5.2 SNA measures

The SNA measure data was taken and put into UCInet for the centrality measures. Since we had the poster and reader data we ran the data to indicate the direction in both ways. So the output shows two sets of centrality measures, Rp and Pr. The Rp is the centrality measures where we ran the UCInet program using the network with directed edges running from the reader to the poster. The Pr is the centrality measures with directed edges running from poster to reader. In our past research, we have found that there was some correlation between hr and Rp Betweenness, Pr Degree, and Pr Betweenness, some negative correlation with Rp Closeness, little correlation with Rp Degree, and no

correlation with Pr Closeness. We also found there was high correlation between hv and Rp Betweenness, Pr Degree, and Pr Betweenness, high negative correlation with Rp Closeness, some correlation with Rp Degree, and no correlation with Pr Closeness (citation withheld for review). Nonetheless we decided to keep all the measures since the current study is exploratory in nature and we do not know which measures might allow us to find about the community members of BigU.

8.5.3 Aggregate measure

Since the SNA had six measures and h-index had two, we summed the SNA measures and divided by six, while we summed the h-index measures and divided by two to get the 'Rank Sum'. We wanted to see which members of BigU would be close to the top of the rankings of the aggregate measure. So table 8-1 is the first page of ordering the 'Rank Sum' measure in ascending order.

8.6. Discussion and Implications

We see from the top 30 with 'Rank Sum' there are very little lurkers in the data (refer to table 8-1 in the appendix). The rankings of the h-indices shows all are in the top 100. The typical high end user using our aggregate measure is a contributor who has more readership than reading other posts. This is shown by the Pr Degree number being higher than the Rp Degree number. On average the Pr degree to Rp degree ratio is two to one (1015 PrDegree to 504 RpDegree). That is the average high user has 2 people reading their posts for each post the high end user read.

We also find that the two Closeness centrality measures seem to have no bearing on the 'Rank Sum'. The average rankings are also problematic with Closeness as they are fairly high (Pr 810 and Rp 2174). During the timeframe of the data collection there were approximately 2200 users that had gotten a rank for the h-indices and there were approximately 2600 users that posted or read a post (for SNA). The seemingly negligible results for the closeness measure may lead us to remove it from the aggregate measures in the future.

We also see that the Betweenness measures are the same for Rp and Pr. This is expected as someone that is between in a directed graph in one direction will be the intermediary for the connection from user A to user B. For the other direction they will be the intermediary for the connection from user B to user A. So the same number appears for betweenness. While one might make an argument for removing betweenness centrality from our measure we feel that looking at the ranking at an average of 163, this measure may hold some more information.

Looking at the hr and hv indices we see that the hv index is much higher than the hr index. This is expected since the views are much easier to attain than replies to posts. What is interesting is to see if there are people that have a high ratio of hr/hv where there is high reply compared to views. There are two users that stand out. User 2523 and user 246 with hr/hv of 0.16 and 0.22 respectively. These two are producing 3 replies per each 20 views and 2 replies per each 10 view respectively. While these users may be praised for their ability to cause interaction they may be individuals that post many controversial messages that get many replies. While those with a low hr/hv ratio (or those that rank high with hv but not so much with hr such as user 549 who ranks 4th in hv but 39th in hr) may be information providers that give relevant information that is read but maybe not commented on so much. These information providers may be more like newspaper article writers where there might be minimal dialog between the author and viewer.

Postings were taken over the whole timeframe of the existence of BigU. So the rankings of those numbers may be skewed as we see many lower level ranks for postings. But we find that these are still in the top 900 (out of 2600). There is one person that has lurking tendencies. That is user 3339, which is identified because of the low rank in h-indices, an hr of 4 for a rank of 76 and hv of 8 for a rank of 89. This user also has very low PrDegree (8) and fairly high RpDeg (768) which tells us that they have only 8 people reading their posts, which is consistent with the low hv and hr indices, and they are reading many users posts (768 of them).

8.6.1 Compared to Old Methods

The use of user status would be given to a user once they post a certain number of posts since the time they started posting in the OC. While this is great to encourage users to come back and post often, we find that these statuses are artificially inflated as we saw in the dataset. Since the measure is a product of postings which is taken over all time the OC has been in existence we feel that a more democratic view of looking at a measure that takes the behavior of other users of the OC would be a much better measure of influence. A dataset such as SNA measures sees the behavior of the readers of the OC. A dataset such as the h-indices measures the behavior of other users in their viewing and replying to other users.

8.6.2 Additional Benefits

In addition we find that the aggregate measure is a much better measure in that it can easily look at a short timestamp of data to find who is relevant in the past short time period, whether it be a few weeks or months. We also feel that using the aggregate measure we can find different categories of

users. We have started to identify some of these such as ‘controversial messenger’, ‘lurker’, or ‘information provider’. We feel that future research can use these types of categories to allow the owner of the OC to target different users for possible behavior changes. Maybe try to get the lurker to contribute more. Maybe see if the ‘controversial messenger’ is creating incendiary posts and target them for possible disciplinary action, such as being banned from the OC. Or maybe encourage the ‘information provider’ to put out more posts that can stir up more discussion.

8.6.3 Implication for Practitioners

We feel that this measure will be a better way to evaluate users for owners of OC’s. The current method that most OC’s use with rating people based solely on the amount of postings or content creation does not give a measure of impact. For practitioners, the interest should be on how impactful certain users posts are. This means that a measure like ours should be used for encouraging users that participate.

Another implication is that as seen in our demonstration, different types of users start to emerge using our measure. Users such as ‘controversial messenger’, ‘lurker’, or ‘information provider’ start to emerge. This allow the practitioner to identify these different types of users and target encouragement depending on the type of users and how the practitioners want the users to improve the vibrancy of the OC.

This also allows a total analysis of the OC with the possibility of using archival data on the OC. Compared to using survey methods, which can bias the output due to response trends in surveys, using the non-invasive measures allows all data to be collected.

8.6.4 Limitations

While the measures identified were an improvement over past methods, there are some limitations. First the dataset used is not generalizable to other OC’s. The current data was taken on a free, university sports fan OC. The generalizability not only to other university fan OC’s or other sports fan OC’s but to other types of OC’s comes into question. A sports fan OC also has the advantage of having topics of conversation that gets updated every season and game-to-game. For another type of OC such as a computer gaming OC, once the community members finish playing the game and conversations about these games have been saturated, and there is minimal topics of conversation to continue the OC. For a sports fan OC this is different as many possible topics of conversation can be planted from game to game. New players might be recruited, new injuries might occur, and game strategies can be discussed from game to game. So the generalizability of the dataset can be

questioned in several ways. In addition we used a six-month period and were able to get a very large dataset. The type of OC to generalize to might be different from an OC that might not have this type of activity week to week.

Another limitation may be the use of replies and views. As the hr and hv indices have been derived from the h-index, the replies and views for hr and hv, respectively, were substituted for the citations counted in the h-index. One can argue that replies to an OC thread is nothing compared to the work that goes into a citation. Worse yet, the views are even easier for an OC member than the replies. One can in addition argue that the hv might be less important, since the views do not necessarily mean that one has 'viewed' the message, it only means that the message was clicked on and displayed on the computer. Another issue that arises from views being easier than replies is that since the views are easier the hv-index is considerably higher than the hr-index. So while the hv-index might mean less due to the ease of views over replies, the hv-index may have larger numbers, which makes comparing the two more difficult.

A third set of limitations comes from the use of the h-index methods in our measures. The h-index has been criticized for several items. Identifying all criticism and presenting them is not in the scope of the current research. The main problems cited with the h-index are the fact that the h-index is harder to gain for more junior researchers, the h-index does not measure recent research, and the h-index does not take into account of the large citation hits. The first two criticisms of the h-index are intertwined. One fact about the h-index is that it can never go down. This is due to the fact that once you gain a citation, that citation never disappears. So if a researcher gets an h-index of 1, that number will never go back down to zero. This ties to the fact that a junior researcher has a harder time garnering an h-index than a more seasoned researcher, since recent publications have a difficult time to gain an h-index. The h-index is tied with citations and for recent publications; it takes time to garner those citations. One cannot predict the citation trend down the road so it becomes harder for junior researcher to gain an h-index.

The third issue with the h-index is that the papers that garner a large number citations, the "one hit wonders" tend to be overlooked by the h-index. The h-index is a measure of both publication productivity and the impact of those publications via citations. If there is a researcher that published two papers over their career but those two papers received thousands of citations, the h-index for that researcher will not be able to get higher than two. Due to the definition of the h-index, which counts the number of publications, the maximum h-index that one can get is limited by the number of publications that the researcher has published. In the same vein for the hr and hv index, are limited to the number of thread starts that an OC member has begun.

8.7. Conclusion

We commenced this research to find new methods for OC owners to find users that contribute to the vibrancy of the OC. The problem that OC's face is the lifecycle of the OC and the possibility of losing relevancy and timely information and becoming a statistic in the graveyard of past OC's. We propose an aggregate measure using SNA and bibliometric measures that are dependent on the behavior of the OC members. These measures are shown to be more descriptive and insightful than the current status quo of using user status or posting frequencies.

We ran data on BigU an online community for a university sports team. The data was taken over a six-month period and showed the diverse nature of the top contributors to the OC. We were able to identify several categories of users that can be helpful for an owner of the OC to identify high-level users.

Appendix (Table 8-1)

Table 8-1. Data Analysis (Shown from left to right, user ID, then the measures are shown with their rankings for each of the measures, Post count, which was the old measure, Poster to Reader Degree Centrality, Closeness Centrality, Betweenness Centrality, Reader to Poster Degree Centrality, Closeness Centrality, Betweenness Centrality, HReplies, HViews, and Rank Sum. The table has been sorted by Rank Sum in Ascending order)																				
Old Measure		New Proposed Measures																		
		Poster to Reader							Reader to Poster							Bibliometric				
ID	Posts	Rk	PrDeg	Rk	PrClose	Rk	PrBetwness	Rk	RpDeg	Rk	RpClose	Rk	RpBetwness	Rk	HR	Rk	HV	Rk	RkSum	
837	6586	48	859	113	4129622	862	19637.002	30	648	162	745767	2319	19637.002	30	43	17	705	6	597.5	
2544	4850	82	684	140	4129596	890	11673.236	69	657	159	745897	2256	11673.236	69	45	15	489	13	611.1666667	
1099	12112	21	368	255	4129352	1054	11838.781	68	1429	6	746130	2148	11838.781	68	48	12	373	22	616.8333333	
702	3998	103	739	131	4129494	985	16339.903	45	662	158	745836	2285	16339.903	45	33	22	532	11	624.6666667	
556	5159	72	6515	8	4129528	961	25159.119	22	822	92	744980	2518	25159.119	22	14	46	360	24	638.8333333	
1032	1417	324	512	187	4129454	1015	5048.833	143	1119	32	745902	2254	5048.833	143	41	18	484	14	645	
1	7129	43	1556	59	4129827	651	21753.754	25	181	766	745379	2437	21753.754	25	134	2	1235	2	662.5	
194	811	527	446	222	4129660	825	3777.453	189	401	381	746006	2197	3777.453	189	45	14	448	17	682.6666667	
3339	947	462	8	661	4129468	1004	18675.486	35	768	114	747609	1760	18675.486	35	4	76	8	89	684	
2167	2330	192	283	302	4130180	368	9747.434	80	61	1189	746650	1900	9747.434	80	17	37	278	27	685.1666667	
2130	6263	53	399	239	4129699	774	4360.307	167	379	407	745991	2203	4360.307	167	21	34	400	19	686	

200	1864	247	590	164	4129672	809	3232.462	216	398	384	745939	2234	3232.462	216	22	31	593	9	690.5
2523	4000	102	208	363	4129425	1026	2257.8	274	1108	34	746428	1988	2257.8	274	30	26	178	39	692.3333333
1472	5087	74	52	548	4129712	762	18317.201	39	251	603	747109	1808	18317.201	39	6	64	53	67	698.6666667
829	544	692	463	206	4129666	820	3653.041	194	381	403	745864	2272	3653.041	194	28	28	464	15	703
1847	2782	148	455	213	4129656	830	2926.133	227	480	295	745942	2228	2926.133	227	11	53	458	16	704.5
1012	1228	370	936	105	4129890	589	7336.154	108	130	904	745547	2387	7336.154	108	54	10	937	3	706.6666667
549	2395	185	784	123	4129649	837	2615.166	243	415	359	745792	2308	2615.166	243	16	39	752	4	707
4125	2514	174	93	479	4129668	817	2759.065	234	452	323	746799	1860	2759.065	234	6	66	95	53	717.3333333
648	1051	431	154	419	4129643	843	1817.942	310	530	249	746326	2036	1817.942	310	46	13	155	46	724
1376	392	877	700	137	4129833	643	3704.386	192	185	754	745754	2324	3704.386	192	22	32	621	8	727
780	2771	150	217	356	4129481	995	2171.137	279	730	131	746237	2090	2171.137	279	13	47	219	35	729.3333333
1110	3795	110	65	525	4129711	763	2480.459	248	383	401	746792	1862	2480.459	248	6	63	66	62	737
246	596	650	179	381	4129592	894	1871.537	306	479	298	746249	2085	1871.537	306	39	19	180	38	740.1666667
192	4940	77	750	129	4129824	657	2396.401	256	162	812	745740	2328	2396.401	256	55	9	747	5	746.6666667
3211	1467	311	237	335	4129566	915	1425.306	328	551	235	746206	2111	1425.306	328	15	45	238	33	747.6666667
1347	1003	448	70	516	4129691	789	2728.755	235	293	529	746781	1863	2728.755	235	13	48	70	60	748.5
312	1490	307	706	135	4129894	588	2711.251	237	120	933	745778	2314	2711.251	237	30	25	655	7	756.6666667
1734	16193	8	5790	11	4129563	921	77178.68	4	914	74	745616	2362	77178.68	4	7	60	0	328	756.6666667
198	4352	96	5649	12	4130051	433	7876.954	100	32	1377	745049	2509	7876.954	100	111	3	2884	1	757.1666667
Mean	3668	246	1015	249	4129668	810	9915.705	163	504	419	746069	2174	9915.7046	163	33	33	489	36	697.5055

Table 8-1. Data Analysis

Chapitre 9

Conclusion générale (FR)

9.1. introduction

Les communautés en ligne (CEL) sont devenues un moyen pour les groupes d'intérêts de partager des idées, communiquer, de débattre, forger des liens en ligne, coordonner des événements hors ligne, développer une communauté en ligne qui bénéficie d'un sentiment d'appartenance. Ces communautés permettent à des groupes d'intérêts communs de transcender le temps et la géographie : le temps, en gardant d'anciennes discussions en ligne comme références pour une utilisation future, et la géographie, en permettant aux usagers d'accéder à la CEL à partir de n'importe quel endroit qui donne accès à l'internet.

Une des difficultés avec les CEL est qu'elles ont tendance à avoir un cycle de vie limité, où leur apparition et leur croissance initiale sont suivies par une phase de stagnation où les usagers cessent d'afficher des commentaires, ce qui mène la communauté à mourir par manque d'activité (Iriberry et Leroy 2009). Le défi est donc de créer une CEL qui continue à prospérer par la production d'information tenue à jour et de communication pertinente. Ma recherche actuelle tente de répondre à cette préoccupation en créant et évaluant des mesures qui permettront aux usagers et aux propriétaires d'identifier les contributeurs à la CEL qui créent un contenu intéressant dont le « buzz » continue à motiver les usagers à afficher des messages et y répondre.

Cette recherche propose l'utilisation de la mesure bibliométrique de l'indice de Hirsch et analyse des réseaux sociaux pour trouver les contributeurs à fort impact. J'utilise une CEL réelle pour évaluer ces mesures. Les mesures proposées sont plus précises que la mesure actuellement utilisée, le « statut de l'utilisateur », qui est seulement basée sur le nombre de messages affichés par des contributeurs sans mesurer leur impact.

Cette partie est organisée comme suit. Dans la section suivante, je vais considérer l'objet de la recherche en général, puis les théories utilisées et évaluées pour cette recherche. J'examine ensuite

quelle est la contribution faite par ce courant de recherche, la méthodologie utilisée, et les données, y compris une présentation de la CEL utilisée pour établir les données. Les résultats de l'étude sont ensuite présentés, suivis d'un survol des limitations et problèmes de cette recherche. Enfin, je discute dans quelles directions s'orientent les recherches futures.

9.2. Objectif

L'objectif principal de cette recherche est de permettre à une CEL de rester pertinente pendant aussi longtemps que possible. Une CEL semble avoir un cycle de vie limité où son dynamisme décroît finalement car elle ne parvient pas à attirer de nouveaux usagers. L'objectif de la recherche actuelle est soit d'étendre la phase dynamique de la CEL, soit d'éviter complètement sa disparition en trouvant des moyens de prolonger sa pertinence à long terme.

Pour qu'une CEL continue à jouir d'une période dynamique, ses propriétaires ont besoin d'identifier les usagers qui contribuent à sa vitalité. Actuellement les CEL utilisent une mesure appelée « statut de l'utilisateur » qui différencie les usagers par leur fréquence d'affichage (Stewart 2005). Par exemple, un usager qui a affiché de nombreux messages peut obtenir un statut de « vétérane », tandis qu'un nouvel usager avec seulement quelques messages obtient une désignation de « débutant ». En indiquant la fréquence des messages, ces désignations mettent seulement l'accent sur la productivité et non sur l'impact. Une telle mesure ne précise pas si un message est capable de créer un long *fil de discussion* dans une CEL ou une discussion animée qui consolidera son dynamisme.

La mesure de l'impact dans les CEL est importante car elle permet de trouver, utiliser et évaluer comment les contributeurs peuvent contribuer à son dynamisme. L'objet de ce travail est concentré sur l'évaluation de la mesure de l'impact possible pour une CEL.

9.3. Théorie et contexte

Cette section se penche sur certaines des théories utilisées dans ce domaine de recherche, puis sur la mesure bibliométrique, et l'analyse des réseaux sociaux.

9.3.1 Théorie du développement de groupe

La théorie du développement de groupe (Back 1951) propose des concepts concernant les liens et l'identité communs dans les groupes. Cette théorie explique comment quelqu'un extérieur à un groupe peut vouloir en devenir membre pour trois raisons. La première, désignée du terme « lien commun », est l'attraction personnelle entre les membres. La deuxième, appelée « identité commune », correspond à l'attraction d'une personne pour l'activité du groupe. La dernière est liée au prestige même du groupe.

Appliqué à une CEL, le « lien commun » décrit la situation où un membre de la CEL attire un membre qui est extérieur, par exemple grâce à la recommandation d'un ami. L'« identité commune » émerge quand une personne est attirée par l'activité que la CEL représente. Normalement, je trouve les usagers de la CEL attirés par une activité commune, et je m'attends donc à voir l'« identité commune » comme raison plus fréquente pour expliquer l'adhésion d'un nouvel usager à une CEL. Enfin, l'attrait représenté par le groupe lui-même n'est peut-être pas aussi répandu dans les CEL, car l'appartenance à la CEL n'est généralement pas attirante pour de nouveaux usagers.

9.3.2 Théorie du capital social

Selon la théorie du capital social, les usagers sont plus susceptibles d'échanger du capital intellectuel et créer un nouveau capital intellectuel lorsque les dimensions structurelles, cognitives et relationnelles sont remplies (Nahapiet et al. 1998). La théorie du capital social a trois dimensions : premièrement, il y a une dimension structurelle où les usagers sont unis par un réseau de liens ; deuxièmement, il y a une dimension cognitive où les usagers partagent un récit et une langue ; enfin, il y a une dimension relationnelle où les usagers ont de fortes relations, telles que la confiance et la coopération, l'obligation, et l'identification.

La dimension structurelle consiste en connexions de réseaux sociaux réalisés dans la CEL. Elle peut être définie par la communication par messages et emails, ou des contributions aux forums, des réponses et les *files* de discussion. La dimension cognitive est représentée par l'intérêt partagé qui motive les usagers à se joindre à une CEL. Enfin, la dimension relationnelle entre en jeu lorsque les usagers trouvent la confiance, les obligations, l'identification, et les normes de la CEL.

9.3.3 Théorie de la présence sociale

Dans la théorie de la présence sociale, la présence est définie comme « l'illusion perceptive de non médiation » (Lombard et al. 1997). Le terme « perceptive » signifie que « ce phénomène implique des réponses continues (temps réel) des systèmes sensoriels, cognitifs, affectifs humains à des objets et entités dans l'environnement d'une personne » (Lombard et al. 1997). « L'illusion de non-médiation » a lieu « lorsqu'une personne ne perçoit ou ne reconnaît pas l'existence d'un intermédiaire dans son environnement de communication et répond comme si elle cet intermédiaire n'était pas là » (Lombard et al. 1997). Cette théorie suggère que le modèle mental substantiel d'un usager « est activé immédiatement après la détection d'un comportement qui suggère la présence d'une autre intelligence » où l'interaction sociale est rendue possible par la technologie (Biocca et al. 2003).

Dans une CEL, cela se voit lorsque les usagers se livrent à une discussion dans un forum. Dans la mesure où les usagers perçoivent les autres usagers comme des êtres intelligents par la communication ou les réponses aux messages, ils commencent à communiquer dans la CEL avec « l'illusion de non-médiation ».

9.3.4 Théorie d'identification sociale

La théorie de l'identification sociale prévoit l'identification d'un usager avec une CEL « que l'on a choisie volontairement résult[ant] d'une appréciation que l'adhésion entraîne des avantages significatifs » (Dholakia et al. 2004). Dans une CEL, cela se passe quand les usagers obtiennent des réponses à leurs questions car il y a un avantage fondamental à recevoir ces réponses. Un autre avantage pourrait être de recueillir de nouvelles informations ou des rumeurs. Les usagers peuvent bénéficier d'informations d'actualité ou profiter des avantages sociaux que leurs donne leur connaissance de rumeurs ou leur participation à une discussion.

9.3.5 Attachement préférentiel

L'attachement préférentiel est d'abord apparu mathématiquement en 1923 avec le modèle des urnes par Eggenberger et Polya (cité par Barbasi et Albert 1999). Le modèle mathématique a été apparu à plusieurs reprises, mais en 1968, Robert Merton l'a appelé l' « Effet Matthieu ». L' « Effet Matthieu » reconnaît « l'immense mérite accordé aux contributions scientifiques de chercheurs de renommée considérable alors qu'elle est refusée aux scientifiques qui n'ont pas encore laissé leur

marque sur un domaine » (Merton 1968). Merton a observé que les coauteurs d'articles scientifiques ne disposaient pas d'une réputation égale et que l'auteur à la plus grande réputation obtenait plus de reconnaissance que les coauteurs de moindre réputation.

La notion d'attachement préférentiel s'applique à l'analyse des réseaux. « Un nœud qui se joint à un réseau, que ce soit une nouvelle page Web ou une nouvelle protéine, peut en principe se connecter à un nœud préexistant. Cependant, l'attachement préférentiel dicte que son choix ne sera pas tout à fait aléatoire, mais linéairement polarisé par le degré des nœuds préexistants - qui est le nombre de liens que les nœuds ont avec d'autres nœuds » (Barbasi 2012 p. 507). Cela signifie que dans un réseau établi de « nouveaux sommets s'attachent de préférence à des sites qui sont déjà bien connectés » (Barbasi et Albert 1999 p. 509).

L'implication de l'attachement préférentiel pour les CEL est que les nouveaux arrivants dans la communauté en ligne vont se joindre aux membres établis. L'attachement préférentiel « se produit dans un réseau lorsque de nouveaux participants décident de choisir pour leurs interactions ceux qui sont déjà bien connectés plutôt que d'autres partenaires » (Faraj et Johnson 2011 p. 1468).

9.3.6 Bibliométrie - Indice de Hirsch

Dans la bibliométrie, l'indice de Hirsch (indice h) utilise le nombre de citations de toutes les publications par un sujet, afin de mesurer l'impact de ce sujet sur le domaine concerné. L'indice h mesure à la fois la quantité de travail qu'un individu a publié et l'impact de cette production en utilisant des citations par d'autres chercheurs pour mesurer « l'influence » de cet individu. Officiellement, l'indice h est défini comme: « Un scientifique a un indice h si h de N_p articles ont chacun au moins h citations, et les autres ($N_p - h$) articles ont au plus h citations chacun » (Hirsch 2005).

L'index h est chargé de créer un « buzz » dans le domaine bibliométrique en produisant des mesures telles que l'indice g (Egghe 2006) et l'indice h_c (Sidiropoulos et al. 2006). Bien qu'il y ait débat quant à la validité et l'utilité de l'indice h et d'autres indices de la famille h, des domaines tels que les systèmes d'information, la chimie, la physique, et de l'économie utilisent cette mesure pour montrer l'influence des chercheurs (Glanzel 2006; Truex et al. 2009). L'indice h est de plus accepté dans le monde universitaire comme mesure de l'influence des chercheurs.

J'ai trouvé des similarités entre l'affichage dans une CEL et la publication universitaire. Tout d'abord, l'activité d'affichage d'un usager est similaire à une publication par un chercheur. Pour la CEL j'ai

deux mesures semblables à la citation d'un article. Je suis capable de mesurer les « réponses » (équivalentes à une citation) et les « consultations » ou messages lus par un usager individuel. La capacité de mesurer les « consultations » est un avantage de la CEL que je ne vois pas dans la bibliométrie. Bien qu'il soit impossible de voir si un chercheur a « lu » l'article d'un autre chercheur, avec la CEL, on peut voir si l'affichage a été « consulté » ou ouvert pour être lu. Donc, je suis en mesure de trouver deux types d'indices h pour la CEL.

J'ai appelé ces deux nouvelles mesures l'indice h_r (réponses) et l'indice h_v (consultations). Pour calculer l'indice h_r il faut d'abord prendre chaque usager dans la CEL et classer ses *fils de discussion* en commençant par le nombre de réponses que chaque *fil* a suscitées. L'usager dispose d'un indice h_r des h_r quand il a au moins h_r *fil* avec des réponses h_r et les *threads* restants ont h_r-1 de réponses ou moins. Par exemple, si un usager a un indice h_r de 100, cela signifie que l'usager a commencé au moins 100 *fil* qui ont obtenu au moins 100 réponses chacun. Donc, cela est un indice de la productivité et de l'impact par le biais des réponses. Le processus d'obtention de l'indice h_v est le même, sauf que les *fils* sont triés et comptés à l'aide des consultations au lieu des réponses.

9.3.7 Analyse du réseau social (ARS)

La centralité dans un réseau social se réfère à la notion que le groupe tourne autour d'un individu. Si un individu est central, plus de personnes sont liées avec lui; s'il est moins central, le nombre de liens à cet individu est moindre. Il y a plusieurs mesures de centralité, et je me sers de trois: degré, proximité, et intermédiarité. Le degré de centralité est un simple comptage des liens qu'un individu possède avec d'autres individus dans le réseau. Le degré se réfère au nombre d'arêtes d'un nœud dans le réseau, ou le nombre de connexions qu'un individu a avec les autres dans le réseau. En théorie, un individu peut avoir au plus $n-1$ connexions quand il y a n individus dans le réseau. C'est parce que l'individu ne peut pas avoir de lien avec lui-même, mais peut avoir des liens avec tout le monde dans le réseau social. Quand un individu a chaque lien possible, la distance entre cette personne et toutes les autres personnes dans le réseau social est 1 (Acedo et al. 2006; Eaton et al. 1999; Henry et al. 2007; Otto et al. 2002; Polites et al. 2008; Vidgen et al. 2007).

La centralité de proximité mesure la proximité d'un individu au reste du réseau. La distance entre deux individus est mesurée par le nombre de liens entre les deux individus prenant le lien le plus étroit entre les deux. Par exemple, si je connaissais un homme (personne A) qui connaît un homme (personne B), puis la distance entre moi et A est 1 et la distance entre moi et B est 2. La centralité de

proximité considère les distances minimales entre un individu et tous les autres membres du réseau et donne la moyenne de ce chiffre (Polites et al. 2008; Vidgen et al. 2007).

Enfin, la centralité d'intermédiarité mesure de la façon dont un individu est utilisé dans les liaisons de distance minimale entre deux autres individus dans le réseau. La centralité d'intermédiarité d'un individu est trouvé en additionnant toutes les distances les plus courtes qui passent par un individu et en divisant par $(n-2)(n-1)$ où n est le nombre de personnes distinctes dans le réseau social (Vidgen et al. 2007).

9.4. Contribution et répercussions

Ma recherche contribue à la recherche sur les CEL en donnant une meilleure mesure des usagers qui contribuent à leur vitalité. L'utilisation de l'indice h et de l'ARS procure une meilleure mesure de l'impact des contributeurs sur le dynamisme des CEL. Ceci représente une nette amélioration par rapport au critère « statut de l'utilisateur » qui mesure seulement la production. Ces nouvelles mesures tiennent compte de l'impact en utilisant les consultations et les réponses aux messages.

Cette recherche étend également au delà du domaine universitaire l'utilisation de mesures bibliométriques, telle la famille de mesures basées sur l'indice h . L'utilisation de cet indice dans les CEL est un prolongement naturel de son utilisation en bibliométrie. Il y a un parallèle entre l'activité de publication universitaire et les contributions aux CEL, car les usagers répondent aux contributions comme les auteurs d'articles de recherche citent d'autres publications.

En tant que contribution universitaire, notre recherche a un impact en aidant à comprendre comment fonctionnent les CEL. D'un point de vue plus pratique, elle a des répercussions pour les propriétaires de CEL car ses résultats leur donnent des outils pour tenter de prolonger le dynamisme de leur CEL. Ce courant de recherche devrait les aider à identifier les contributions à fort impact. Avec cette information, ils peuvent à leur tour encourager les usagers à contribuer davantage de messages dont la pertinence aide à maintenir le dynamisme et la durabilité des CEL.

9.5. Données

Pour la majorité de cette étude, les données viennent BigUFans.com. Un des collègues et coauteurs de cette recherche étant son propriétaire, il a permis l'accès gratuit aux données de cette CEL.

BigUFans.com est une CEL pour l'université Big University. Big U. est une grande institution de recherche de niveau I dans le sud-est des États-Unis. C'est une université publique, fondée il y a 150 ans, avec plus de 30.000 étudiants. La CEL est orientée vers les équipes sportives de Big U. où le football américain est le sport dominant. Big U. participe à la *NCAA (National Collegiate Athletic Association) Division I* dans de nombreuses disciplines sportives, y compris les sports majeurs que sont le football américain, le basketball et le baseball.

9.5. Méthodologie

Les données ont été collectées dans la CEL BigUFans.com. L'utilisation du « statut de l'utilisateur » de BigUFans.com a rendu les données bibliométriques et ARS disponibles. Les données ont été initialement collectées en utilisant un Excel. Une fois les données collectées, l'utilisation d'un programme Java créé par l'auteur a été nécessaire pour calculer les indices h_r et h_v . Pour l'analyse du programme ARS, le logiciel open source UCINET a été utilisé. L'analyse de corrélation a été exécutée avec le logiciel open source R ou SPSS.

La collecte des données a été effectuée à divers moments correspondant aux dates de publication des divers chapitres. La plupart des données ont été collectées sur une période d'un ou deux ans. Les données historiques ont été conservées sur BigUFans.com mais il y avait quelques limitations dues à la taille du dépôt de données. L'analyse des données a été difficile, car leur quantité était énorme, avec, par exemple, certaines données annuelles occupant plusieurs centaines de milliers de lignes d'un fichier Excel.

9.7. Résultats

A ce jour cette recherche a produit cinq résultats préliminaires importants. Tout d'abord, il n'y avait pas de différence dans la quantité de trafic d'affichage des CEL entre les événements importants et anodins liés à au sujet de la CEL. Des événements conséquents (un match contre une équipe rivale) produisent la même quantité de trafic sur la CEL que des faits plus anodins (le recrutement d'un nouveau joueur).

Deuxièmement, les indices h étaient meilleurs que son statut pour prédire l'impact de l'utilisateur. Ceci est prévisible, puisque que seule la fréquence des contributions, et non le statut de l'utilisateur, joue

aucun rôle pour mesurer l'impact des contributions. Troisièmement, dans l'une des études pluriannuelles (sur deux ans) j'ai trouvé que, sur une période d'un an, le « statut de l'utilisateur » suffisait pour prédire sa contribution, mais que sur une période de deux ans, les indices h donnaient une meilleure prédiction de cette contribution.

Quatrièmement, concernant les ARS, j'ai montré une forte corrélation entre la centralité de degré et la centralité d'intermédiarité avec les indices h , mais pas avec la centralité de proximité. Les usagers qui sont proches de tout le monde dans le réseau social n'ont pas nécessairement un grand impact, tandis que ceux qui ont de nombreux liens avec d'autres usagers (centralité de degré) et ceux qui sont des intermédiaires vitaux entre deux autres personnes qui se lient (centralité d'intermédiarité) exercent une plus grande influence.

Enfin pour les nouveaux usagers, je n'ai pas observé d'attachement préférentiel. Les nouveaux usagers étaient plus à l'aise pour démarrer de nouvelles discussions que des usagers plus expérimentés. En outre, les usagers expérimentés avaient tendance à répondre davantage aux autres usagers chevronnés tandis que les nouveaux usagers n'avaient pas cette attitude.

9.8. Limitation

Cette recherche a plusieurs limitations. Premièrement, j'ai utilisé l'indice h , qui mesure à la fois la productivité et l'impact et par rapport au « statut de l'utilisateur » qui est seulement une mesure de la productivité. Les mesures de l'indice h prédiront mieux les usagers qui ont plus d'impact. L'indice h dans la bibliométrie a également été l'objet de quelques critiques et ces limitations ont également cours dans notre recherche. Les principales limites sont que l'indice h nécessite du temps pour que les citations d'auteurs apparaissent et que les certains auteurs, dont les publications sont rares mais très influentes, comme un bestseller isolé, ne recevront pas un score élevé.

Deuxièmement, j'ai utilisé les indices h_r et h_v pour tenir compte des réponses et des consultations. Les « consultations » sont un chiffre beaucoup plus facile à augmenter que les « réponses » et seront donc plus nombreuses. Bien que ma capacité à mesurer les « consultations » soit un avantage sur la bibliométrie (il n'y a aucun moyen de mesurer le nombre de « consultations » d'un article), l'augmentation rapide des chiffres de l'indice h_v comparés aux chiffres h_r n'a pas été étudiée et exige une analyse plus approfondie. Cette limitation peut également venir du fait que l'indice h a été créé pour des articles universitaires, et que les citations sont beaucoup plus difficiles à produire qu'une « réponse » sur un site de CEL.

Troisièmement, j'utilisé une seule CEL pour cette étude (CEL pour sports universitaires). Les résultats peuvent ne pas être généralisables à d'autres CEL et les données peuvent être biaisées par le type de CEL. En quatrième lieu, les données longitudinales ont été collectées seulement pour une brève période de temps, avec un maximum de deux ans. Cela était dû à la grande quantité de données collectées et les limites imposées par le traitement informatique.

9.9. Futures recherches

Compte tenu des limites indiquées dans les sections précédentes, je peux facilement identifier certains domaines pour de futures études. Je dois d'abord utiliser d'autres mesures qui permettent de mesurer l'impact. Des études sont également nécessaires dans l'analyse des indices hr et hv pour examiner comment ils diffèrent des mesures bibliométriques initiales de l'indice h.

Je dois ensuite développer l'étude d'autres CEL, et pas seulement la CEL BigUFans.com. Une étude plus généralisable ou la réplique de la même étude doit se faire en utilisant d'autres plateformes de CEL. Enfin je dois mener une étude longitudinale de plus longue durée pour vérifier que les résultats sont valables à long terme. Je prévois aussi des études qui analysent la CEL actuelle avec un ensemble de données collectées sur une plus longue durée.

Chapitre 10

General Conclusions (EN)

10.1. Introduction

Online communities (OC) have become a means for common interest groups to share ideas, communicate, debate, forge online connections, coordinate off-line events, grow a social community online and a sense of being part of a OC. These communities allow common interest groups to communicate across time, via keeping old threads online for future use and reference, and geography, via allowing users to access the OC online from any location that has access to the internet.

One challenge for the owners of the OC and for contributors that are vested with their time and effort in posting to the OC is that OC's seem to have a life-cycle, where eventually they become irrelevant and die off (Iriberry and Leroy 2009). The challenge here is to create an OC that continues to thrive with relevant, up-to-date information and communication. The current research stream tried to address this concern by creating and evaluating metrics that will allow users and owners to identify those contributors to the OC that are creating content that is creating a 'buzz' and continuing to attract users to view and reply to the posts.

This research proposed the use of the bibliometric measure of the Hirsch-index and social network analysis to find the high impact contributors to the OC. I use a real life OC to evaluate the measures. The proposed measures are more accurate than the currently used measure, the 'user-status' driven only by the number of posts made by the contributors which has no measure of impact.

This chapter is organized as follows. In the next section I look at the object of the overall research, then the background and the relevant theories used and evaluated for use by this research. I then look at the contribution of this research stream, the methodology used, and the data, including a summary of the OC used for data taking. The results of the study is then presented, followed by an extensive look at the limitations and issues with the research. Finally, I discuss what direction future research is heading.

10.2. Objective

The main objective of this research stream is to allow OC's the ability to stay relevant for as long as possible. The vibrancy of an OC seems to have an inevitable life-cycle where it eventually falls out of interest by the community and fails to attract new users. The current research objective is to either extend the OC in the vibrant era during the life-cycle, or completely avoid the end by continuing to find ways for an OC to continue to be relevant for an extended period of time.

For an OC to continue, the vibrancy period the owners need to identify those users who are contributing to the vibrancy of the OC. Currently OC's use a 'user-status' measure which differentiates users by their frequency of posting (Stewart 2005). For example, a user that has posted many posts may get an 'all-star' user-status, while a new user with only a few posts might get a 'rookie' user-status designation. While these user-status designations indicate frequency of posts, the designation is only an indication of productivity and not of impact. The current user-status measure does not indicate if a poster is able to create a long thread in an OC or a lively discussion, which would add to the vibrancy of the OC.

One aspect that is not seen is some measure of impact in OC's so to find, use, and evaluate measures that can assess impact and thus, contributors that can create vibrancy in the OC is important for this field. The object of this work is concentrated on the evaluation of possible impact measure for an OC.

10.3. Theory and Background

This section will look at some of the theories used in this research stream, then the bibliometric measure, and social network analysis.

10.3.1 Group Development Theory

Group development theory (Back 1951) had some concepts about common bond and common identity in groups. This theory explains how someone outside of a group would want to join in a group. There were three reasons why someone would want to join a group. The first is personal attraction with members named 'common bond'. The second is when a person is attracted to the group activity named 'common identity'. The final one is when there is a prestige to the group itself.

In OC terms 'common bond' would apply when there is a member in the OC that someone on the outside is attracted to. This might be because of recommendation by a friend. For 'common identity' the person is attracted to the activity that the OC is representing. Normally I find users of OC's attracted by some common activity so I expect to see 'common identity' to explain new user sign up

in an OC more prevalently than the other attractions. Finally the group itself being attractive may not be as prevalent in OC's, as being in the OC is usually not an attraction for new users.

10.3.2 Social Capital Theory

According to social capital theory, users are more likely to exchange intellectual capital and create new intellectual capital when structural, cognitive, and relational dimensions are met (Nahapiet et al. 1998). Social Capital Theory has three dimensions. First is the structural dimension where users are connected through a network of ties. The second is the cognitive dimension where users share a narrative and language. Finally there is the relational dimension is where users have strong relational characteristics such as trust, norms of cooperation, obligation, and identification.

In OC's, the structural dimension is the social network connections made in the OC. The structural dimension in OC's can be defined by communication via messages and emails, or forum contributions and the replies and discussion threads in the OC. The cognitive dimension is the shared interest that brings users to the OC. Finally the relational dimension comes into play when users find trust, obligations, identification, and norms of the OC.

10.3.3 Social Presence Theory

In social presence theory, presence is defined as "the perceptual illusion of nonmediation" (Lombard et al. 1997). The term 'perceptual' means "this phenomenon involves continuous (real time) responses of the human sensory, cognitive, and affective processing systems to objects and entities in a person's environment" (Lombard et al. 1997). The 'illusion of nonmediation' happens "when a person fails to perceive or acknowledge the existence of a medium in his/her communication environment and responds as he/she would if the medium were not there" (Lombard et al. 1997). This theory suggests a substantial mental model of a user "is activated immediately upon detection of behavior that suggests the presence of another intelligence," where the social interaction is mediated by technology (Biocca et al. 2003).

In an OC, this is seen when users engage in discussion in the forum. As users sense other users as an intelligent being via communication or replies to posts, they start to communicate in the OC with the 'illusion of nonmediation'.

10.3.4 Social Identification Theory

Social identification theory predicts a user's identification with an OC “that one has chosen volitionally stems from an understanding that membership entails significant benefits” (Dholakia et al. 2004). In an OC, I see this as users get responses to their questions. There is an inherent benefit to having questions answered. Another benefit might be garnering new information or gossip. Users can benefit from getting up-to-date information or may be getting the social benefits of gossiping or being in a discussion.

10.3.5 Preferential Attachment

Preferential attachment first mathematically appeared in 1923 with the urn model by Eggenberger and Polya (as cited by Barabasi and Albert 1999). The mathematical model had been repeatedly appeared but the model was termed the “Matthew Effect” by Robert Merton in 1968. The “Matthew Effect” is where “the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark” (Merton 1968). Merton observed that co-authors of scientific papers did not have equal reputation and that the author with the higher reputation got more recognition than the less reputable co-authors.

The preferential attachment notion has translated into network analysis. “A node joining a network, such as a new web page or a new protein, can in principle connect to any pre-existing node. However, preferential attachment dictates that its choice will not be entirely random, but linearly biased by the degree of the pre-existing nodes – that is, the number of links that the nodes have with other nodes” (Barabasi 2012 p. 507). This means that in an established network “new vertices attach preferentially to sites that are already well connected” (Barabasi and Albert 1999 p. 509).

The implication of preferential attachment for OC's is that new entrants into the online community will attach themselves to established community members. Preferential attachment “occurs in a network when new actors choose to interact with already well-connected others over more typical others” (Faraj and Johnson 2011 p. 1468).

10.3.6 Bibliometrics – Hirsch Index

In bibliometrics, the Hirsch index (h-index) utilized citation counts of all the publications by a subject to measure the impact of that subject on the field. The h-index measures both the amount of work that a subject has published as well as the impact of that body of work using citations by other researchers to measure the ‘influence’ of the subject. Formally the h-index is defined as: “A scientist

has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each" (Hirsch 2005).

The h -index is responsible for creating a buzz in the bibliometric field spawning other measures such as the g -index (Egghe 2006) and the hc -index (Sidiropoulos et al. 2006). While there is debate as to the validity and usefulness of the h -index and other h -family of indices, areas such as information systems, chemistry, physics, and economics have used this measure to show influence of the researcher (Glanzel 2006; Truex et al. 2009). The h -index is becoming accepted in academics as a measure of influence.

I find similarities to posting in an OC and academic publishing. First, a user posting is similar to a publication by a researcher. For the OC I have two measures that are similar to the citation of a paper. I am able to measure the 'replies' similar to the citation and I am also able to measure the 'views' or the posting being read by an individual user. The ability to measure the 'views' is an advantage of the OC that I do not see in bibliometrics. While it is impossible to see if a researcher has 'read' another researchers article with the OC, I can see if the posting has been 'viewed' or in essence opened to be read. So I am able to find two types of h -indices for the OC.

I termed the two new measures as the hr -index (replies) and the hv -index (views). To compute the hr -index one must first take each user in the OC and rank the user's threads starting with the number of replies each thread has garnered. The user has an hr -index of hr when they have at least hr threads with hr replies and the remaining threads have $hr-1$ replies or less. For example, if a user has an hr -index of 100 then that means the user has started at least 100 threads that have gained at least 100 replies or more each. So this is an index of both productivity and impact via replies. The process for obtaining the hv -index is the same except the threads are sorted and counted using views instead of replies.

10.3.7 Social Network Analysis (SNA)

Centrality in a social network refers to the notion that the group revolves around a certain individual. If an individual is more central, there are more people connected to that individual; if they are less central, there are less connections to that individual. There are several measures of centrality, three of which I use: degree, closeness, and betweenness. Degree centrality is a simple count of the connections an individual has with others in the network. Degree refers to the number of edges a node in the network has, or the number of connections that an individual has with the others in the network. In theory an individual can have at most $n-1$ connections when there are n individuals in the social network. This is because the individual cannot have a connection with themselves, but can

have connections with everyone else in the social network. When one individual has every possible connection, the distance between that individual and all other persons in the social network is one (Acedo et al. 2006; Eaton et al. 1999; Henry et al. 2007; Otto et al. 2002; Polites et al. 2008; Vidgen et al. 2007).

Closeness centrality is a mean measure of how close an individual is to the rest of the network. The distance between two individuals is measured by number of connections between the two individuals taking the closest connection between the two. For example, if I knew a guy (individual A) who knows a guy (individual B), then the distance between me and A is one and the distance between me and B is two. The closeness centrality takes the minimum distances between an individual and all others in the network and gives the mean of this number (Polites et al. 2008; Vidgen et al. 2007).

Finally the betweenness centrality is a measure of how an individual is used in the minimum distance connections between any two other individuals in the network. The betweenness centrality of an individual is found by the summing all the shortest distances that go through an individual and dividing by $(n-2)(n-1)$ where n is the number of distinct individuals in the social network (Vidgen et al. 2007).

10.4. Contribution and Implications

The current research contributes to OC research with a better measure of users that contribute to the vibrancy of the OC. The use of the h-index and SNA is a better measure of the impact contributors have on the OC vibrancy. This is a marked improvement over the current user of 'user status,' which is only a measure of production. The new measures take into account the impact using replies and views of the contributor's posts.

This research also extends the use of bibliometric measures such as the h-index family of measures outside of the academic publishing realm. The use of the h-index in OC's is a natural extension of the use of the bibliometric h-index. There is a similarity between academic publishing and OC contributions as users can reply to OC contributions and authors of research articles can cite other publications.

Academically the impact of this research will help us understand how OC's operate. For practitioners, there are implications as OC owners can use the research results to try to extend the vibrancy of their OC. This research stream should help owners of OC's to identify the high impact contributions.

With this information OC owners can encourage these users to produce more postings that are impactful and help keep the OC vibrant and healthy.

10.5. Data

Data for the majority of this study has been taken from BigUFans.com. One of the colleagues and coauthors of this research stream is the owner of BigUFans.com allowing free access to data on the OC. BigUFans.com is a OC website for the school Big University. Big U. is a large tier I research institution in the southeastern US. Big U. is a public state university that has been around for 150 years with over 30,000 students. The OC is geared towards the athletic teams of Big U. where football is the major sport. Big U. competes in the NCAA (National Collegiate Athletic Association) Division I in many athletic disciplines including the major sports of football, basketball, and baseball.

10.6. Methodology

Data was taken from the BigUFans.com OC. By using BigUFans.com user status, bibliometric, and SNA data was available. Data was initially taken using a text file read into excel or read directly into an excel file. Once data was taken, use of a java program created by the author was needed to calculate the hr and hv indices. For the SNA analysis the open source UCINET program was used. Correlation analysis was run on the open source R program or SPSS.

Data collection was done at various times corresponding to the publication dates of the individual chapter venues. Most data was taken over a one or two year period. Historical data has been kept on BigUFans.com but there were some limitations on repository size. Data analysis has been challenging as the amount of data has been large, for example, some annual data points required several hundreds of thousands of lines of an excel spread sheet.

10.7. Results

The research stream to date has five important preliminary results. First, there were no differences in the amount of OC posting traffic with major and minor events associated with the OC interest. Big events (a game against a rival) and minor events (a signing of a new recruit) produced the same amount of traffic on the OC.

Second, the h-indices were better at predicting user impact than user status. This is expected as user status has no part in measuring impact of contribution, only contributor frequency. Third, in one of the multi year (two year) studies I found that in the one year period the user status was adequate in

predicting user contribution but over a two year period the h-indices were better at predicting user contribution.

Fourth, regarding SNA, there was strong correlation support with degree and betweenness centrality with the h-indices but not for closeness centrality. Those users who are close to everyone else in the social network are not necessarily going to be highly impactful contributors, while those that have many connections (degree centrality) with other users and those that are important intermediaries for two other people to connect (betweenness centrality) are going to be high impact contributors.

Finally for new users, support of the preferential attachment was not seen. New users were comfortable starting new threads more so than veterans. Additionally, veterans tended to reply more to fellow veterans while new users did not have this bias.

10.8. Limitation

With the current stream of research there are several limitations. First, I used the h-index, which measures both productivity and impact and compared to the 'user status,' which is only a measure of productivity. The h-index measures will be better at predicting users that have more impact. The h-index in bibliometrics has also seen some criticism and those limitations also apply here. The major limitations being that the h-index requires time for citations to occur and highly impactful authors that only publish a few works, like the one-hit-wonders, will not score high.

Second I used the hr and hv index to account for reads and views. The 'views' are a much easier number to augment than the 'replies' and will run up higher. While the fact that I am able to measure the 'views' is an advantage over bibliometrics (there is no way to measure how many 'views' an article gets), the fast increase in the hv numbers compare to hr numbers is not studied and requires more analysis. This limitation also can criticize the fact that the h-index was created for academic articles, and that citations are much more difficult to produce than the 'reply' on an OC site.

Third, I used only one OC for this study. The findings may not be generalizable to other OC's and that the data may be biased by the type of OC (college sports fan OC). Fourth, the data was taken for longitudinal data but for only a small period of time, with a maximum of two years. This was due to the large amount of data taken and computing limitations that ensued.

10.9. Future

Given the limitations on the previous sections, I can easily identify some future areas of study. First I need to use other measures that measure impact. Studies are also needed in the analysis of the h_r and h_v indices and how they are different from their bibliometric origin measures of the h -index and if there are differences.

Second the study needs to expand into other OC's, not just the BigUFans.com OC. A more generalizable study or replication of the same study needs to take place using other OC platforms. Finally a longer longitudinal study needs to be conducted to verify that the findings hold over time. Future studies to try to look at a longer set of data with the current OC are in the plans.

References

- Acedo, F.J., Barroso, C., Casanueva, C., and Galan, J.L. "Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis," *Journal of Management Studies* (43:5), July 2006.
- Assmann, J., Sandner, P., and Ahrens, S. "Users' Influence on the Success of Online Communities," in: *Proceedings of the 42nd Hawaii International Conference on System Sciences*, 2009.
- Back, K. "Influence through social communication," *Journal of Abnormal and Social Psychology* (46:1) 1951, pp 9-23.
- Barabasi, Albert-Laszlo, and Reka Albert. "Emergence of Scaling in Random Networks." *Science* 286 (October 15, 1999): 509–12.
- Barabasi, Albert-Laszlo. "Luck or Reason." *Nature* 489 (September 27, 2012): 507–8.
- Biocca, F., Harms, C., and Burgoon, J.K. "Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria," *Presence: Teleoperators & Virtual Environments* (12:5), October 2003 2003, pp p456-480.
- Burke, M., Marlow, C., and Lento, T. "Feed Me: Motivating Newcomer Contribution in Social Networking Sites" in: *Proceedings of the CHI 2009 Social Networking Sites conference*, 2009
- Chen, J., Xu, H., and Whinston, A.B. "Moderated Online Communities and Quality of User-Generated Content," *Journal of Management Information Systems* (28:2), Fall 2011 2011, pp 237-268.
- Cheng, R., and Vassileva, J. 2005. "User Motivation and Persuasion Strategy for Peer-to-Peer Communities," in: *Hawaii International Conference on System Sciences*. Hawaii.
- Cheng, R., and Vassileva, J. "Design and evaluation of an adaptive incentive mechanism for sustained educational online communities," *User Modeling and User-Adapted Interaction* (16:3) 2006, pp 321-348.
- Cheshire, C. and Antin, J. "The Social Psychological Effects of Feedback on the Production of Internet Information Pools," *Journal of Computer-Mediated Communication* (13), 2008, pp 705-727.
- Clavio, G. "Uses and gratifications of Internet collegiate sport message board users," D., INDIANA UNIVERSITY, 2008.
- Clark, J.G., Warren, J., and Au, Y.A. "Carnegie Classifications and Institution Productivity in Information Systems Research: A Scientometric Study," *Communications of the Association for Information Systems* (19) 2007, pp 478-512.
- Collins, U. *Collins English Dictionary: 30th Anniversary Edition*, (10th ed.) HarperCollins, London, 2010.
- Cothrel, J., and Williams, R. "On-line communities: helping them form and grow," *Journal of Knowledge Management* (3:1) 1999, pp 54-60.

- Cuellar, M., Young, B., Takeda, H. (March 2012) "Online Community User Influence: A Study Using User Status" Proceedings of the Southern Association for Information Systems conference (SAIS) March 23-24th, 2012, in Atlanta, GA
- Cuellar, M. Young, B. Sainsbury, R., Takeda, H. (August 2013) "Online Community of Influence: A Study Using the Hirsch Metric and Social Network Analysis" Proceedings of the Americas Conference on Information Systems (AMCIS) August 15-17th, 2013 in Chicago, IL
- Dantzig, P. (July, 2002). "Architecture and design of high volume web sites:(a brief history of IBM sport and event web sites)." Paper presented at the The Fourteenth International Conference on Software Engineering and Knowledge Engineering, Ischia, Italy 2002
- Dholakia, U.M., Bagozzi, R.P., and Pearo, L.K. "A social influence model of consumer participation in network- and small-group-based virtual communities," *International Journal of Research in Marketing* (21:3) 2004, pp Pages 241-263.
- Eaton, J.P., Ward, J.C., Kumar, A., and Reingen, P.H. "Structural Analysis of Co-Author Relationships and Author Productivity in Selected Outlets for Consumer Behavior Research," *Journal of Consumer Psychology* (8:1) 1999, pp 39-59.
- Eggenberger, Von F., and G. Polya. "Über Die Statistik Verketteter Vorgänge." *Journal of Applied Mathematics and Mechanics* 3 (1923): 279–89.
- Egghe, L. "Theory and practice of the g-index," *Scientometrics* (69:1) 2006, pp 131-152.
- Faraj, Samer, and Steven L. Johnson. "Network Exchange Patterns in Online Communities." *Organization Science* 22, no. 6 (December 2011): 1464–80.
- Freeman L. C. "A Set of Measures of Centrality Based on Betweenness" *Sociometry* (40:1) 1977, pp 35-41
- Fuller, J., Bartl, M., Ernst, H., and Muhlbacher, H. 2004. "Community Based Innovation: A Method to Utilize the Innovative Potential of Online Communities," in: Hawaii Internation Conference on System Sciences. Hawaii.
- Gainer, E., Sollet, C., Ulmann, M., Lévy, D., and Ulmann, A. "Surfing on the morning after: analysis of an emergency contraception website." *Contraception*, (67:3) 2003, pp. 195-199.
- Gangi, P.M.D., Wasko, M.M., and Hooker, R.E. 2010. "Getting Customers' Ideas to Work for You: Learning from Dell How to Succeed with Online User Innovation Communities," *MIS Quarterly Executive* (9:4), Dec. 2010, pp. 213-228.
- Ganley, D., and Lampe, C. "How Deep Runs the Karma? Structural Holes and Social Capital in an Online Community," in: Midwest Association for Information Systems Conference, Grand Rapids, MI, 2006.
- Gefen, D., Straub, D., and Boudreau, M.-C. "Structural Equation Modelling and Regression: Guidelines for Research Practice," *Communications of the Association for Information Systems* (4:7) 2000.

- Glanzel, W. "On the h-index--A mathematical approach to a new measure of publication activity and citation impact," *Scientometrics* (67:2) 2006, pp 315-321.
- Golle, P., Leyton-Brown, K., Mironov, I., and Lillibridge, M. (2001). "Incentives for sharing in peer-to-peer networks." *Electronic Commerce*, 75-87.
- Gómez, V., Kaltenbrunner A., and V. López. "Statistical analysis of the social network and discussion threads in Slashdot," in *Proceedings from WWW conference*, page 645, New York, New York, USA, Apr. 2008. ACM Press
- Harper, F., Frankowski, D., Drenner, S., Ren, Y., Kiesler, S., Terveen, L., et al. (2007). Talk amongst yourselves: inviting users to participate in online conversations. Paper presented at the International Conference on Information Systems, Montreal
- Henry, N., Goodell, H., Elmqvist, N., and Fekete, J.-D. "20 Years of Four HCI Conferences: A Visual Exploration," *International Journal of Human-Computer Interaction* (23:3) 2007, pp 239-285.
- Henseler, J., Hubona, G., and Ringle, C.M. "Structural Equation Modeling Using Smart PLS," in: *Training class on using SmartPLS*, Atlanta, GA, 2008.
- Hirsch, J.E. "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences of the United States of America* (102:46) 2005, pp 16569-16572.
- Hirsch, J.E. "Does the h index have predictive power?," *Proceedings of the National Academy of Sciences of the United States of America* (104:49), December, 4 2007, pp 19193-19198.
- Holmstöm, H., Henfridsson, O. "Improving Packaged Software through Online Community Knowledge," *Scandinavian Journal of Information Systems* (18:1) 2006
- Howison, James; Wiggins, Andrea; and Crowston, Kevin (2011) "Validity Issues in the Use of Social Network Analysis with Digital Trace Data," *Journal of the Association for Information Systems* : Vol. 12: Iss. 12, Article 2. Available at: <http://aisel.aisnet.org/jais/vol12/iss12/2>
- Iriberry, A., and Leroy, G. "A life-cycle perspective on online community success," *ACM Comput. Surv.* (41:2), February 2009, p 29 pages.
- Johnson, Steven L., Samer Faraj, and Srinivas Kudaravalli. "Emergence of Power Laws in Online Communities: The Role of Social Mechanisms and Preferential Attachment." *MIS Quarterly* 38, no. 3 (September 2014): 796–808.
- Jung, J. H., Schneider, C., and Valacich, J. "Enhancing the Motivational Affordance of Information Systems: The Effects of Real-Time Performance Feedback and Goal Setting in Group Collaboration Environments," *Management Science* (56:4) 2010, pp 724-742
- Kankanhalli, A., Tan, B. C. Y., Wei, K. "Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation," *Management Information Systems Quarterly* (29:1) 2005, pp. 113-143
- Keen, P. "MIS research: reference disciplines and a cumulative tradition." 1980

- Koh, J., Kim, Y., Butler, B., and Bock, G. "Encouraging participation in virtual communities." *Communications of the ACM*, (50:2) 2007, 73.
- Lampel, J., and Bhalla, A. "The role of status seeking in online communities: Giving the gift of experience," *Journal of Computer-Mediated Communication* (12:2) 2007, pp 434-455.
- Laniado, D., Tasso, R., Volkovich, Y., Kaltenbrunner, A. "When the Wikipedians Talks: Network and Tree Structure of Wikipedia Discussion Pages," *Proceedings of ICWSM 2011*
- Law, S.P.-M., and Chang, M.K. "Fostering Knowledge Exchange in Online Communities: A Social Capital Building Approach," in: *International Conference on Information Systems, Paris, France, 2008*.
- Lombard, M., and Ditton, T. "At the Heart of It All: The Concept of Presence," *Journal of Computer Mediated Communication* (3:2) 1997.
- Ma, M. and Agarwal, R. "Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities," *Information Systems Research* (18:1) 2014, pp. 42-67
- Marett, K., and Joshi, K.D. "The Decision to Share Information and Rumors: Examining the Role of Motivation in an Online Discussion Forum," *Communications of the Association for Information Systems* (24:4) 2009, pp 47-67.
- Merriam-Webster Online Dictionary. "Lurking". Retrieved from <http://www.merriam-webster.com/dictionary/lurking> 2010
- Merton, Robert K. "The Matthew Effect in Science." *Science* 159, no. 3810 (January 5, 1968): 56–63.
- Mousavidin, E., and Goel, L. "A Life Cycle Model of Virtual Communities," in: *Proceedings of the 42nd Hawaii International Conference on System Sciences, 2009*.
- Nahapiet, J., and Ghoshal, S. "Social Capital, Intellectual Capital, and the Organizational Advantage," *The Academy of Management Review* (23:2), April 1998, pp. 242-266.
- Otto, E., and Rousseau, R. "Social Network Analysis: A powerful strategy, also for the information sciences," *Journal of Information Science* (28:441) 2002.
- Polites, G.L., and Watson, R.T. "The Centrality and Prestige of CACM," *Communications of the ACM* (51:1), January 2008, pp 95-100.
- Porter, C. "A typology of virtual communities: a multi-disciplinary foundation for future research," *Journal of Computer-Mediated Communication* (10:1) 2004.
- Postmes, T., and Spears, R. Refining the cognitive redefinition of the group: Deindividuation effects in common bond vs. common identity groups, 2000, pp. 63-78.
- Preece, J. "Sociability and usability in online communities: determining and measuring success." *Behaviour & Information Technology*, (20:5) 2001, pp. 347-356.

- Preece, J. "The top five reasons for lurking: improving community experiences for everyone." *Computers in Human Behavior*, (20:2) 2004, pp. 201-223.
- Rashid, A. M., Ling, K., Tassone, R. D., Resnick, P., Kraut, R., and Riedl, J. (April 22-27, 2006). "Motivating Participation by Displaying the Value of Contribution." Paper presented at the CHI, Montreal, Quebec Canada. 2006
- Ren, Y., and Kraut, R. (2007a). "An Agent-Based Model To Understand Tradeoffs In Online Community Design." Paper presented at the International Conference on Information Systems.
- Ren, Y., Kraut, R., and Kiesler, S. "Applying common identity and bond theory to design of online communities," *Organization Studies* (28:3) 2007b, p 377.
- Ren, Y., Harper, F. M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., Kraut, R. E., "Building Member Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds" *Management Information Systems Quarterly* (36:3) 2012 pp. 841-864
- Ringle, C.M., Wende, S., and Will, A. "SmartPLS," University of Hamburg, Hamburg, Germany, 2005, p. Software to perform PLS analyses.
- Rothaermel, F., and Sugiyama, S. "Virtual internet communities and commercial success: Individual and community-level theory grounded in the atypical case of TimeZone. com." *Journal of Management*, (27:3) 2001, 297.
- Sassenberg, K., and Postmes, T. (2002). "Cognitive and strategic processes in small groups: Effects of anonymity of the self and anonymity of the group on social influence." *British Journal of Social Psychology*, 41(3), 463-480.
- Schaedel, U., and Clement, M. "Managing the Online Crowd: Motivations for Engagement in User-Generated Content," *Journal of Media Business Studies* (7:3) 2010, pp 17-36.
- Shen, K.N., and Khalifa, M. "Design for Social Presence in Online Communities: A Multi-Dimensional Approach," in: Pacific Asia Conference on Information Systems, Suzhou, China, 2008.
- Shen, K.N., and Khalifa, M. "Design for Social Presence in Online Communities: A Multidimensional Approach," *Transactions on Human-Computer Interaction* (1:2), June 2009, pp 33-54.
- Shen, K.N., Khalifa, M., and Yu, A.Y. "Supporting Social Interaction in Virtual Communities: Role of Social Presence," in: Americas Conference on Information Systems, Acapulco, Mexico, 2006.
- Shiue, Y., Chiu, C., and Chang, C. "Exploring and mitigating social loafing in online communities." *Computers in Human Behavior*, (26:4) 2010, pp. 768-777.
- Shrout, P.E., and Bolger, N. "Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations," *Psychological Methods* (7:4) 2002, pp 422-445.
- Sidiropoulos, A., Katsaros, D., and Manolopoulos, Y. "Generalized h-index for Disclosing Latent Facts in Citation Networks," arXiv:cs.DL/o606066 (1), 07/13/2006 2006.

- Skretta, D. "Recruiting Web sites booming, also under fire." from Retrieved April 15, 2010, from <http://nbcsports.msnbc.com/id/17973156/> 2007
- Spring, S., and Zheng, W. "Reputation Systems of Online Communities: Establishing a Research Agenda," in: Midwest Association for Information Systems Conference, Eau Claire, WI, 2008.
- Stewart, D. "Social Status in an Open-Source Community," *American Sociological Review* (70:October), October 2005 2005, pp 823-842.
- Takeda, H., Cuellar, M., Young, B., Sainsbury, R. (August 2012) "What Drives User Contribution in an Online Community? A Study in Contributor Influence and User Status" Proceedings of the Americas Conference on Information Systems (AMCIS) August 9-11th, 2012 in Seattle, WA
- Takeda, H. (2010) "Examining scholarly influence: a study in Hirsch metrics and social network analysis" Dissertation, Georgia State University.
- Trevino, L. K. "Business Ethics and the Social Sciences." In R. E. Frederick (Ed.), *A Companion to Business Ethics* (pp. 218-230). 1999 Oxford, England: Blackwell Publishers.
- Truex III, D.P., Cuellar, M.J., and Takeda, H. "Assessing Scholarly Influence: Using the Hirsch Indices to Reframe the Discourse," *Journal of the Association of Information Systems* (10:7) 2009, pp 560-594.
- Truex III, D.P., Cuellar, M.J., Takeda, H., and Vidgen, R. "The Scholarly Influence of Heinz Klein: Ideational and Social Measures of His Impact on IS Research and IS Scholars," *European Journal of Information Systems* (20:4) 2011.
- Vidgen, R., Henneberg, S., and Naude, P. "What sort of community is the European Conference on Information Systems? A social network analysis 1993-2005," *European Journal of Information Systems* (16) 2007, pp 5-19.
- von Krogh, G., Haefliger, S., Spaeth, S. and Wallin, M. W. "Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development," *Management Information Systems Quarterly* (36:2) 2012, pp. 649-676
- von Krogh, G., Spaeth, S. and Lakhani, K. R. "Community, joining, and specialization in open source software innovation: a case study," *Research Policy* (32) 2003, pp. 1217-1241
- Wang, Y., and Fesenmaier, D. (2003). "Assessing motivation of contribution in online communities: An empirical investigation of an online travel community." *Electronic Markets*, 13(1), 33-45.
- Wang, Y., and Fesenmaier, D. R. (2004). "Towards understanding members' general participation in and active contribution to an online travel community," *Tourism Management*, 25, 709-722.
- Wang, Y., Liu, Y., Zhang, M., Ma, S., and Ru, L. (2008, January 7-12). "Identify Temporal Websites Based on User Behavior Analysis." Paper presented at the Third International Joint Conference on Natural Language Processing, Hyderabad, India.

- Wasko, M. M. and Faraj, S. "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *Management Information Systems Quarterly* (29:1) 2005, pp. 35-57
- Wilson, S. M., and Peterson, L. C. "The Anthropology of Online Communities." *Annual Review Anthropology*, 31, 2002, pp. 449-467.
- Young, B., Takeda, H., & Cuellar, M. "Investigating the Impact of Offline Events on Group Development in an Online Sports Community," in: *Southern Association for Information Systems (SAIS)*, Atlanta, GA, 2011a.
- Young, B., Takeda, H., Cuellar, M. "All Contributions are Not Created Equal: Measuring User Influence in Online Communities with Hirsch Indices," in: *International Federation For Information Processing Working Group 8.2 Organizations and Society in Information Systems Workshop (IFIP 8.2 OASIS)* Shanghai, China, 2011b.
- Zhang, Z. "Feeling the Sense of Community in Social Networking Usage," *IEEE Transactions on Engineering Management* (57:2) 2010, pp 225-239.
- Zheng, J., and Kim, H.-W. "Investigating Knowledge Contribution from the Online Identity Perspective," in: *International Conference on Information Systems*, Montreal, Quebec, 2007.

Table des matières

Remerciements	3
Sommaire	6
Chapitre 1	7
Introduction générale (FR)	7
1.2.1. Communautés en ligne (CEL).....	10
1.2.2 Méthodes invasives	12
1.2.3 Bibliométrie	13
1.2.4 Analyse du réseau Social	15
1.3. Contexte de la recherche	16
1.4. Présentations à des conférences.....	18
1.4.1 Étude de l'impact des événements hors ligne sur le développement du groupe dans une communauté sportive en ligne.....	18
1.4.2 Influence de l'usager d'une communauté en ligne: étude avec User Status.....	18
1.4.3 Quels facteurs déterminent la contribution des usagers dans une communauté en ligne? Etude sur l'influence des contributeurs et sur le statut des usagers	19
1.4.4 Influence d'une communauté en ligne: étude utilisant l'analyse métrique de réseau social de Hirsch.....	19
1.4.5 Mesurer la contribution d'un nouveau membre dans une communauté en ligne: approche par l'attachement préférentiel.....	20
1.4.6 Mesurer l'impact des contribution des membres d'une communauté en ligne: une nouvelle approche inspirée par l'indice de Hirsch	20
1.5. Analyse et répercussions.....	21
1.5.1 Contributions.....	21
1.5.2 Répercussions pour les professionnels	22
1.5.3 Limites	22
1.6. Tableau de synthèse des recherches	24
1.7. Conclusion	27

Chapitre 2	28
General Introduction (EN)	28
2.1. Introduction	29
2.2. Background.....	30
2.2.1 Online Communities.....	31
2.2.2. Invasive Methods	33
2.2.3. Bibliometrics.....	34
2.2.4. Social Network Analysis.....	36
2.3. Research Context.....	37
2.4. Papers	38
2.4.1. Investigating the Impact of Offline Events on Group Development in an Online Sports Community	38
2.4.2. Online Community User Influence: A Study Using User Status.....	38
2.4.3. What Drives User Contribution in an Online Community? A Study in Contributor Influence and User Status	39
2.4.4. Online Community of Influence: A Study Using the Hirsch Metric and Social Network Analysis.....	39
2.4.5. Measuring New Member Contribution in an Online Community: A Preferential Attachment Approach	40
2.4.6. Measuring Member Contribution Impact in an Online Community: A New Hirsch Index Inspired Approach.	40
2.5. Discussion and Implications	41
2.5.1. Benefits.....	41
2.5.2. Implication for Practitioners.....	41
2.5.3. Limitations	42
2.6. Table of Research	43
2.7. Conclusion	46
Chapitre 3	47
Investigating the Impact of Offline Events on Group Development in an Online Sports Community..	47

Abstract	47
3.1. Introduction.....	48
3.2. Literature Review	48
3.3. Research Questions.....	49
3.4. The Empirical Context	50
3.5. Research Model and Variables of Interest	51
3.6. Hypotheses.....	53
3.7. Preliminary Analysis	53
3.8. Conclusion	55
Chapitre 4.....	57
Online Community User Influence: A Study Using User Status.....	57
Abstract	57
4.1. Introduction.....	58
4.2. Literature Review	58
4.3. Research Context.....	60
4.4. Measures	61
4.5. Research Proposal	62
4.6. Summary.....	63
4.7. Conclusion	64
Chapitre 5.....	66
What Drives User Contribution in an Online Community ? A Study in Contributor Influence and User Status.....	66
Abstract	66
5.1. Introduction.....	67
5.2. Background.....	68
5.2.1. Assessing OC's based on User Status	68
5.2.2. Assessing Vibrancy of the OC	69
5.3. Theoretical Framework and Hypothesis Development.....	70

5.4. Research Context.....	71
5.5 Methodology	73
5.5.1. Measures	73
5.5.2. Instrument Validation.....	74
5.6. Findings and Hypotheses.....	76
5.7. Summary.....	76
5.8. Conclusion	78
Chapitre 6	79
Online Community Influence: A Study Using the Hirsch Metric and Social Network Analysis	79
Abstract	79
6.1. Introduction.....	80
6.2. Background.....	81
6.2.1. Assessing OC's based on User Status	82
6.2.2. Assessing Vibrancy of the OC	82
6.2.3. Social Network Analysis.....	84
6.3. Theoretical Framework and hypothesis development	84
6.4. Research context.....	85
6.5. Methodology	86
6.6. Measures	87
6.7. Findings and hypotheses	87
6.8. Summary.....	88
6.9. Conclusion	90
Chapitre 7	91
Measuring New Member Contribution in an Online Community: A Preferential Attachment Approach	91
Abstract	91
7.1. Introduction.....	92
7.2. Background.....	93

7.2.1. Online Communities.....	93
7.2.2. Preferential Attachment.....	96
7.3. Propositions.....	97
7.4. Research Context.....	97
7.5. Research Methodology	99
7.6. Results	99
7.7. Discussion and Implications	100
7.8. Conclusion	102
Chapitre 8	103
Measuring Member Contribution Impact in an Online Community: A New Hirsch Index Inspired Approach	103
Abstract	103
8.1. Introduction	104
8.2. Background	106
8.2.1 Online Communities	106
8.2.2 Invasive Methods.....	108
8.2.3 Bibliometrics	109
8.2.4 Social Network Analysis.....	111
8.3. Research Context.....	112
8.4. Research Methodology.....	113
8.5. Measures.....	114
8.5.1 h-index measures.....	114
8.5.2 SNA measures	114
8.5.3 Aggregate measure.....	115
8.6. Discussion and Implications.....	115
8.6.1 Compared to Old Methods	116
8.6.2 Additional Benefits	116
8.6.3 Implication for Practitioners	117

8.6.4	Limitations	117
8.7.	Conclusion.....	119
Chapitre 9	122
Conclusion générale (FR).....		122
9.1.	introduction.....	122
9.2.	Objectif.....	123
9.3.	Théorie et contexte	123
9.3.1	Théorie du développement de groupe	123
9.3.2	Théorie du capital social.....	124
9.3.3	Théorie de la présence sociale	124
9.3.4	Théorie d'identification sociale	125
9.3.5	Attachement préférentiel	125
9.3.6	Bibliométrie - Indice de Hirsch	126
9.3.7	Analyse du réseau social (ARS).....	127
9.4.	Contribution et répercussions.....	128
9.5.	Données.....	128
9.5.	Méthodologie	129
9.7.	Résultats	129
9.8.	Limitation.....	130
9.9.	Futures recherches.....	131
Chapitre 10	132
General Conclusions (EN)		132
10.1.	Introduction.....	132
10.2.	Objective.....	132
10.3.	Theory and Background.....	133
10.3.1	Group Development Theory.....	133
10.3.2	Social Capital Theory	134

10.3.3 Social Presence Theory.....	134
10.3.4 Social Identification Theory.....	134
10.3.5 Preferential Attachment.....	135
10.3.6 Bibliometrics – Hirsch Index.....	135
10.3.7 Social Network Analysis (SNA).....	136
10.4. Contribution and Implications.....	137
10.5. Data	138
10.6. Methodology	138
10.7. Results	138
10.8. Limitation.....	139
10.9. Future	139
References.....	141
Table des matières	148
Table des figures.....	155
Table des tableaux.....	156

Table des figures

Figure 3-1. The Conceptual Model	51
Figure 3-2. Model Analysis Results.....	55

Table des tableaux

Table 1-1. Tableau de Recherche	25
Table 2-1. Table of Research	44
Table 3-1. Results of Bootstrapping Analysis	55
Table 4-1. User Status Breakdown	61
Table 5-1. BigUFans.com user status/posting required and coding of StatusNo.....	72
Table 5-2 Correlation Table of Predictor variables vs. output measures using Kendall's Tau	75
Table 6-1. The Correlation Matrix	89
Table 7-1. Count of Posts and Ratio	100
Table 7-2. First posts by new user and ratio	100
Table 7-3. Average of thread starter number of posts for replies	100
Table 8-1. Data Analysis	121