Predicting Internet-based Online Community Size and Time to Peak Membership Using the Bass Model of New Product Growth

David R. Firth, Cameron Lawrence, and Shawn F. Clouse University of Montana, Missoula, MT, USA

david.firth@business.umt.edu cameron.lawrence@business.umt.edu shawn.clouse@business.umt.edu

Abstract

The objective of this paper is to use a well-established theory of new product diffusion and show how it can be applied to cyber communities to forecast how many people will join the community, and how long it will take to attain peak membership. We use three online communities to illustrate the application of the theory, and show how the theory can be used to measure community size and time to peak membership for these communities. We show that the model can be used to delineate the two different types of influence (internal and external) that impact joining the communities. The paper is appropriate for researchers wishing to better understand the mechanics underlying online community growth, and for administrators of such communities who want to forecast the important aspects of size and peak adoption time.

Keywords: Online community, Bass Model, technological innovation, institutional processes

Introduction

Online communities are "social aggregations that emerge from the Internet when enough people carry on...public discussions...with sufficient human feeling to form webs of personal relationships in cyberspace" (Reingold, 1993, p.5). The notion of a cyber community has been at the heart of the Internet since its inception and, if used in a skillful way, it seems likely that they will help build new and deeper relationships with customers (Armstrong & Hagel, 1996). Such relationships can then form the basis for fruitful social and commercial interaction.

Although their current use is low, increased use of online communities by merchants and the entertainment industry shows that they are an important area for businesses to understand. Recently, television shows have created online communities to allow viewers to interact with each other after the show. AOL (AOL, 2000) reported that its most popular web site was for one such show, with more than one message being posted per second to a specially created online community.

In creating an online community, a community administrator starts with nothing except the technology in place to enable communications and to provide a memory of the communication that will occur. The community then builds from there and can be considered an *emergent* community. Despite the fact that online communities have been around for over a decade, we

Material published as part of this journal, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Publisher@InformingScience.org to request redistribution permission.

were able to find little research addressing the growth of such emergent communities. This paper seeks to take a first step in filling that gap. By examining three emergent online communities at a major U.S. university's business school, we show how the administrators of such communities can predict the size and time to peak adoption of the community. Both are important measures as they give a sense of how the community is developing. Armed with this knowledge, skillful administrators can choose interventions that might impact the development of the community.

The paper proceeds with a discussion of a theory for online community growth. We then present our research setting, and use the theory to predict the eventual number of users of each community studied. We conclude with implications for research and practice.

Theory, Research, Findings, and Discussion

Theory

In this paper we propose that Internet-based online communities are an example of an innovation. As people join the community, the innovation diffuses. The diffusion of an innovation can be defined as the process by which that innovation is communicated through certain channels over time among the members of a social system (Rogers, 1983). The innovation diffuses as people learn about it through various channels, such as the Internet, or through face-to-face interactions. Similarly, online communities diffuse as people come to hear about them over the Internet or from people they meet.

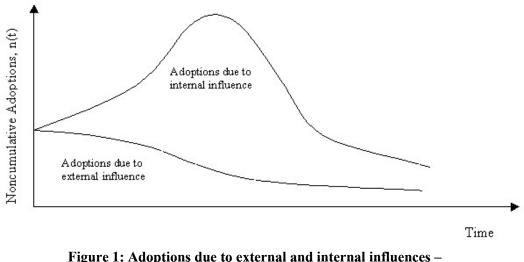
In looking for a theory to explain the diffusion of an Internet-based online community innovation, the *Bass new product growth model* for consumer durables (Bass, 1969) appears as a good candidate. To our knowledge, this is the first time the Bass Model has been used to examine online communities. As a model of diffusion, it has formed the basis of a great deal of diffusion of innovation research in the marketing field (Mahajan, Mueller, & Bass, 1990). The model centers around an important notion of a "first-purchase sale", which means a sale is one that the "purchase volume per buyer is one unit" (Mahajan et al., 1990, p.2). In looking at emergent online communities, the notion translates very well - a "first-purchase sale" suggests that a person adopts the community once only. In essence, the purchase volume per buyer is one unit, that unit being the community.

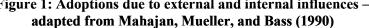
In many contexts, the purchase of a particular good is rather clear cut, and can be measured at the point either a commitment to exchange funds is established, or when actual funds change hands. What then is the equivalent of time of purchase in an online community? Prior research into online communities has shown that certain members may never actually take part in the community. Called "lurkers" (Nonnecke & Preece, 2001) these people observe what transpires within the online community, but never actually contribute to the community by posting a message that may be consumed by others. We define lurkers as individuals who have not "purchased" the community – they are like window shoppers at a mall. This, therefore, means we define purchase in an online community as one who posts a message to the community. Posting a message suggests that a person has bought-in to the community. For many online community administrators this is a useful definition, as lurkers often are hard to track and, potentially, add little value to the community. In contrast, members of the community who post a message can be analyzed from those messages (Firth, 2002), and can provide value to other members of the community.

The basic premise of the *Bass new product growth model* (Bass, 1969) is that part of the influence affecting adoption depends on imitation, and part of it does not (Mahajan et al., 1990). These two influences are commonly termed external influence and internal influence (Lekvall & Wahlbin, 1973; Mahajan et al., 1990). Using the Bass Model to model online community adoption, we

posit that the diffusion of an innovation such as an Internet online community is therefore due to a combination of internal and external influences, which we now explain in further detail.

Internal influence implies that online communities are subject to the contagion effect such that diffusion occurs through interpersonal contacts. This means that the interaction of prior adopters with potential adopters affects the rate of adoption. This occurs as people use the community, see the ways that the community is being used, and how many people are in the community. As more people adopt, potential adopters see the increased value brought about by the increased network, and adopt themselves. Under external influence, adopters are not influenced by people who have already adopted, and are generally thought to be influenced by mass media. That is, they see external sources, such as magazines, news reports and advertisements, and it is these factors that induce them to adopt the community. Figure 1 shows non-cumulative adoptions as a result of interpersonal (internal) and mass media (external) communication channels. The model shows that those adopting as a result of mass media (external) are present throughout the process, and are brought on board by the mass media in decreasing amounts as time passes. Those adopting through the internal interpersonal mode follow a traditional peaked path, reflecting the notion that such adoption requires a build-up of people before it takes hold, reaches a peak, and then subsidies as the potential adoption population becomes saturated. Looked at cumulatively, adoptions due to internal influence would have a classic s-shaped curve.





In the Bass model, coefficients are assigned to the internal and external influences, with p representing the coefficient of external influence and q representing the coefficient of internal influence. The Bass model is based on a hazard function, which represents the probability that an adoption will occur at time t, given that it has not yet occurred. With n(t) being the number of adopters at time t, N(t) being the cumulative number of adopters at time t, m representing the potential number of ultimate adopters, and p and q representing the coefficients of external and internal influence, the Bass model can be represented (more detail on this formula can be found in Bass, 1967; Mahajan et al., 1990) as:

$$N(t) = p[m - N(t)] + (q/m) N(t)[m - N(t)]$$
(1)

With a minimum of three points available for estimation of the parameters, this equation reduces to:

$$n(t+1) = \alpha_1 + \alpha_2 N(t) + \alpha_3 N^2(t)$$
 (2)

where $\alpha_1 = pm$, $\alpha_2 = q - p$, and $\alpha_3 = -q/m$.

In our context, N(t) and n(t) are known by capturing the time that an individual commenced using an online community in an active way by posting a message. From this, an OLS estimation of the α 's in equation 1 then allows us to recover estimates of p, q and m.

In addition to being able to estimate values of p, q and m, from the Bass model equation (equation 1) we can differentiate to find the time that adoptions reach their peak. This is given by:

$$T(\text{peak}) = 1/(p+q)Ln(q/p)$$
(3)

Equations 1 and Equation 3 are most relevant to us here, as they allow us to predict Internet-based online community size and the time it takes to achieve peak adoption of the community. In the following research study, we use these equations to show how we can estimate online community size and how long a community takes to reach its peak size, and then we compare these estimates to actual results.

Research Study

The context of this research is a set of emergent online communities for admitted students at a major U.S. university's business school. After offering admittance, the business school administration provides the student access to a host of resources to assist in the student's transition to attending business school. One of the main elements of this assistance is a website that has links to academic issues, administration issues, and an online community. Along with their offer letter, a student is sent information about this website because it provides a useful way for students to serve themselves when it comes to resolving the many, mainly repetitive (for the administration) issues that arise. The online community is featured in that letter, as the administration sees the community as a way to foster a sense of "school spirit" amongst the new students, even before they arrive on campus.

All admitted students had access to the community over the Internet, evidenced by each student having their own school-assigned e-mail address. Three types of students are admitted; full time MBAs in a two-year program (MBA), fully employed MBAs (FEMBA), who take the program evenings, one weekday and one weekend day for three years, and executive MBAs (EMBA) who take the program one weekday and one weekend day for two years. The EMBA community, a small community due to the limited program size, is not studied here. The technology and space for each community was provided by the administration of the MBA programs, and they developed a web site dedicated to incoming students, with a link to each program, which in turn had a link to the appropriate online community. As such, the online community is open to all who know its URL (Uniform Resource Locator or internet location). A username and password is required, but a person may select any name and any password so there is no real limitation on access. The MBA online community commenced Saturday, March 18, 2000. The FEMBA online community commenced Thursday, May 25, 2000.

In addition, on Monday, May 1, 2000, the San Francisco Bay Area MBAs formed their own online Internet community at www.egroups.com, which we will call BAEG (Bay Area eGroup). Just as with the MBA and FEMBA communities, BAEG is open to all who know its URL. Al-

though a username and password is required, once again a person may select any name and any password to gain access.

We note that access to the three communities is almost identical, removing one element of contextual variability between the communities. This reduces the possibility that patterns observed in the communities are an artifact of the processes by which users can set themselves up in a community. Specific contextual differences do exist:

The *MBAs* are in general a geographically dispersed community - 24% come from outside the United States, and 93% from outside the town where the university is located.

The *FEMBAs* are in general a geographically proximate community. The vast majority (>95%) live within 2 hours drive of the university.

The *BAEG* is a geographically proximate sub-set of the MBA community, with members collocated in the Bay Area of San Francisco, providing them with the opportunity to meet in person. Indeed part of the community's stated purpose was "to plan summer mixers".

	MBA	FEMBA	BAEG
Launch date of the community	March 18, 2000	May 25, 2000	May 1, 2000
Demographics	24% non-U.S. 93% non-Los Angeles	4% non-Los Angeles 96% within 2 hrs of Los Angeles	98% within the Bay Area of San Fran- cisco
People posting messages	191 out of 327	65 out of 190	58 out of 58
Messages posted	582	296	355
Primary topic(s)	Moving to and living in Los Angeles (31% of messages)	Getting to know each other online (31% of messages)	Moving to and living in Los Angeles (36% of messages)
			Getting together in the Bay Area (35% of messages)

Table 1: Profile of the Communities

Table 1 shows descriptive details for each community. On a daily basis we collected information about who posted a message to the community. Postings to all three communities ceased on or before the start of term, Monday October 4th, 2000 providing a natural end date for data collection for the study. The primary data for the study consisted of 1,233 messages posted by the newly admitted students; 582 messages in the MBA online community posted by 191 students, 296 messages in the FEMBA by 65 students, and 355 messages in the BAEG by 43 students.

The communities studied here are of a particular class - members of the community are *admitted* to it as part of a marshalling-in process. In these communities, individuals meet online first, and then physically meet later. This class of community is a prevalent form: new hires at a large organization are part of an admitted group; the formation of groups for a new project at a company begins with the admittance of team members; promoted individuals in an organization are admitted to a new echelon of the organization.

Findings

Given that N(t) (the cumulative number of adoptions at time t) and n(t) (the number of adoptions at time t) are known by capturing the time that an individual commenced using an online community, equation 2 can be used to estimate the values of p, q and m, where p and q are the coefficients of external and internal influence, and m is the potential number of ultimate adopters. These estimates are shown in Table 2 below. Since we know the actual number of people who adopted the communities, we report this as well for comparison.

Community	Р	q	т	Actual <i>m</i>
MBA	0.017	0.238	192	191
FEMBA	0.057	0.288	66	65
BAEG	0.175	0.000	49	43

Table 2: Estimates of p, q and m for the MBA, FEMBA and BAEG communities

Using these coefficients in the Bass model (equation 1) above, we can plot the mathematical representation of the diffusion process for each community – Figures 2, 3, and 4. Since we also know the actual profile of the communities, we plot this also.

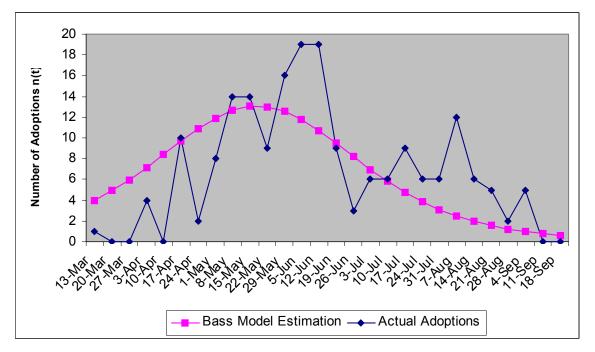


Figure 2: Plot of the Bass Model Estimation and Actual Adoptions for MBAs

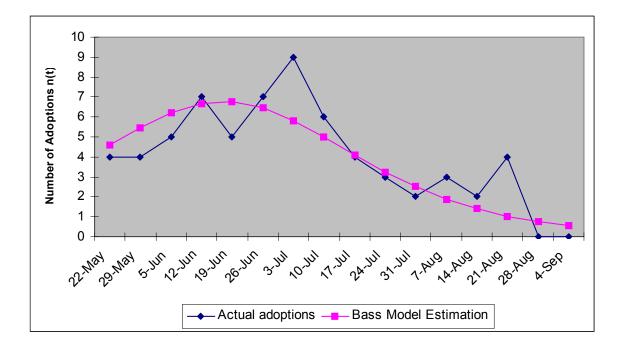


Figure 3: Plot of the Bass Model Estimation and Actual Adoptions for FEMBAs

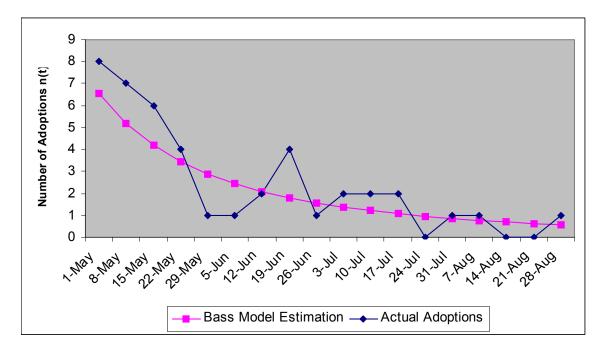


Figure 4: Plot of the Bass Model Estimation and Actual Adoptions for BAEG

Since we have estimates of p and q for each community, we can use equation 3 to predict the time that adoptions in the community will reach their peak. It should be noted that we set time period 1 for the actual adoptions as that period where adoptions equal or exceed pm for the first time (Bass, 1969), as the Bass model has pm as a solution when t=0. The predicted versus actual time

of peak adoption is shown in Table 3, and in each community studied, the actual as compared to predicted time of peak adoption is quite close.

Table 3: Comparison of the predicted and actual time of peak adoption for the MBA, FEMBA and BAEG communities

Community	Predicted	Actual
MBA	10.42	9
FEMBA	4.72	4
BAEG	1	1

To this point we have been comparing the predictive ability of the Bass model to the actual data based on a complete data set – that is we have been looking retrospectively at the data. It is also possible to see how well the Bass model will perform in predicting n(t) and m (the total number of adopters) from incomplete data. That is, how well does the Bass model forecast n(t) and m as time progresses. As already noted, with a minimum of three points, the values of p, q and m can be estimated. This means it is possible to derive values for N(t) and n(t) for

a range of values of t, from t > 3 up to t_{max} that is the value of t when use of the community ended (t_{max} MBA = 29, FEMBA = 16 and BAEG = 18).

Table 4: Estimates of *m* (the total number of adopters), and the sum of squared differences between the actual and estimated curves for the MBA, FEMBA and BAEG communities across differing values of t

actoss uniciting values of t				
Community	Estimated m	Sum of		
	(the total number	Squared		
	of adopters)	Differences		
MBA	(actual m = 191)			
t=5	7	2167.11		
t=11	112	2806.17		
t=15	180	588.84		
t=20	183	590.85		
t=25	208	481.69		
FEMBA	(actual m = 65)			
t=5	43	154.49		
t=10	72	41.64		
BAEG	(actual m = 43)			
t=5	27	39.13		
t=10	43	24.11		
(t=10 MBA is an unstable solution and generates invalid results)				

In Table 4 we use the Bass model (equation 1) to estimate *m* (the total number of adopters) for a selection of t from t > 3 up to t_{max} for each community. In addition, we can derive an estimate of the adoption curve using this equation and equation 2. To check the accuracy of the estimate, at each incremental value of t the sum of the squared differences between the actual and estimated curves can be calculated. A closer fit of the estimate to the actual curve is suggested by a lower sum of squared differences.

As expected, the estimated value of *m* (the total number of adopters) gets closer to the actual value as the number of points used to estimate increases to the maximum. We also see that additional data enables us to reduce the sum of squared distances between the actual and predicted curves. A non-parametric sign test can be used to compare the sum of the squared differences at each t to the sum of the squared differences at t_{max}. The sign test shows us that in the MBA community, only t=25 is statistically close to the estimate using all data points (t_{max}=29). For the FEMBA community, t=10 is statistically close to the estimate using all

data points, and for the BAEG, t=5 provides a statistically good estimate.

Given the equation for time to the peak period of adoptions (equation 3), it is also possible to estimate the time of peak adoptions using a range of values of t, from t > 3 up to t_{max} . This is shown in Table 5.

Table 5: Estimates of the time to the
peak period of adoptions for the MBA,
FEMBA and BAEG communities
given limited data

8		
Community	Estimated	
	time	
MBA (actual time = 10)		
t=5	1	
t=11	3	
t=15	9	
t=20	9	
t=25	10	
FEMBA (actual time = 4)		
t=5	3	
t=10	5	
t=15	5	
BAEG (actual time = 1)		
$t=5$ to t_{max}	1	
(t=10 MBA is an unstable solution and		
generates invalid results, and the period		
of maximum adopters for the BAEG is		
set at 1 given that $p \ge q$ based on the		
shape of the curve.)		

In the MBA community, the estimated peak time for adoptions is estimated exactly at t=25, although t=15gives a very close prediction. For the FEMBA community, t=5 and t=10 both give close estimates. Given that the BAEG has a coefficient of internal influence of zero, all estimation curves predict maximum adoptions at t=1, which is the actual time of maximum adoptions.

Discussion

Our results show how effective the Bass model is in predicting the number of adopters of an online community as well as the time to peak adoption. The estimates of the values of p and q in Table 1, as well as plots of the curves in Figures 2, 3 and 4, show clearly the difference between how the BAEG community, and the MBA and FEMBA communities diffuse. With bigger values of q than p, both the MBA and FEMBA communities reflect adoptions due to internal influence. With a bigger value of p than q (in fact q= 0) it is clear that the BAEG community reflects adoptions due to external influence.

Our understanding of how these three communities emerged supports this finding. The MBA and FEMBA communities were established ahead of

potential members knowing of their existence. Potential members were informed of the communities by e-mail, and membership grew as more and more people experienced the communities. The utility derived from the community increases correspondingly with an increase in the set of users with whom a member wishes to communicate (Gurbaxani, 1990), through internal influence.

The BAEG had different foundations. The initial members of this community were already members of the MBA community. Further, the benefits of community membership were already known, and so when the BAEG was launched and advertised in the MBA community, those in the bay area already using the MBA community moved en-mass over to using the BAEG community. From a BAEG community perspective, this represents an external event, and can be described as external influence. It is also likely, however, that members of the BAEG were subject to internal influence in the MBA community. The strong evidence of external influence in Figure 4 suggests, though, that the majority of influence in the BAEG is external.

The values of m (the total number of adopters) calculated from the Bass model using all available data (Table 1) are highly congruent with the actual value of *m*. Table 2 also shows that the model is able to accurately estimate when the peak adoption period will occur. However, such estimates are of limited practical value since they are derived from a complete set of data when the actual total number of adoptions is already known. What these results do show is the viability of using the Bass equations to model adoptions in an online community.

Of value to practitioners is the ability to predict the total number of adopters of the community, and the period that peak adoption occurs. Table 3 shows that for online communities subject to internal influence (MBA and FEMBA), reasonable estimates of the total number of adopters, m, comes after the peak of noncumulative adoptions, as suggested by Heeler and Hustad (1980) and

Srinivasan and Mason (1986). For all communities, a reasonable estimate of m (within 80% of the final actual value) is obtained using $2/3^{rd}$ of all available data.

The time to peak adoption is considered to be the strategically most interesting in marketing (Mahajan et al., 1990), as managers can use this information to make judgments about future expenditures made in connection with the product. In an emergent online community the time to peak adoptions is of interest as it shows the point that growth of the community is beginning to slow. Community administrators can use this information to target advertising to new potential members of the community, or to introduce content that might stimulate lurkers to be actual community members. Our results (Table 4) show that the peak point can be detected within a rather short time after the actual peak – for the MBA community by t = 15 (out of a total of 28) we can be reasonably accurate about the time of the peak ($t_{estimated peak} = 9$ as opposed $t_{actual peak} = 10$), and for the FEMBA community by t = 5 (out of a total of 18). For the BAEG we know almost immediately that the peak has been reached.

Conclusion

It is important to note that this research was conducted at one site, and as such we may have issues in being able to generalize our results to other settings. However, the people adopting the communities were not merely students, although they are fairly unique in that they have all elected to study for an MBA. The average work tenure of the MBA and BAEG is 4 years, whereas that of the FEMBA is 7 years. At the time these people adopted an online community, they were still part of the workforce (and for the FEMBAs, would remain so). As such, we feel that the generalizability of these results is strengthened to other commercial (rather than merely educational) communities, although we cannot know this is the case without further research. Indeed, we suggest that future research using the theory laid out would be fruitful in other settings such as online communities specifically dedicated to corporations (e.g. Apple, Microsoft, GE, Ford), specific technologies, entertainment, and such.

Our research is also limited by our definition of an adopter, that is one who posts a message in the community. As has been noted in the Bass Model literature (Mahajan et al., 1990), the model explicitly assumes that each potential adopter joins the community only once. But what happens if a person posts a message, and then doesn't visit the community for a long time, comes back and post another message? Future research should seek to modify the model for such individuals.

We believe this research is a first step in understanding growth in online communities, and in particular how researchers and administrators of such communities can measure and predict the size and time to peak adoption of the community. This is important information as it gives researchers a way to understand community dynamics, and for administrators it tells them when critical points in the lifecycle of the community have been reached.

Acknowledgement

The authors would like to thank Javier Gómez, Assistant Professor of Managerial Economics, at the University of Navarra IESE Business School, and the reviewers. This research was supported in part by the UCLA Anderson School's Information Systems Research Program, and The University of Montana School of Business Administration Summer Research Grant Program.

References

- AOL (America Online). (2000). Retrieved September 24, 2000, from http://www.aol.com
- Armstrong, A. & Hagel, J. (1996). The real value of online communities. *Harvard Business Review*. 74(3), 134-141.
- Bass, F. M. (1969). A new product growth model for consumer durables. *Management Science*. 15(5). 215-227.
- Firth, D. R. (2002) Emergent online communities: The structuring of communicative practices over the Internet. *Proceedings of the 23rd International Conference on Information Systems (ICIS)*. Barcelona, Spain, December 15-18.
- Gurbaxani, V. (1990). Diffusion in computing networks: The case of BITNET. *Communications of the ACM*. 33(12). 65-75.
- Heeler, R. & Hustad, T. P. (1980). Problems in predicting new product growth for consumer durables. *Management Science*, 26, 1007-1020.
- Lekvall, P. & Wahlbin, C. (1973). A study of some assumptions underlying innovation diffusion functions. *Swedish Journal of Economics*, 75, 362-377.
- Mahajan, V., Mueller, E., & Bass, F. M. (1990). New product diffusion models in marketing: A review and directions for research. *Journal of Marketing*, 54(1), 1-16.
- Nonnecke, B., & Preece, J. (2001). Why lurkers lurk. Proceedings of the Sixth Americas Conference on Information Systems. Boston, Massachusetts, August 9-11. 1521-1530.
- Reingold, H. (1993). *The virtual community: Homesteading on the electronic frontier*. Reading, MA: Addison Wesley.
- Rogers, E. M. (1983). Diffusion of innovations (3rd ed.). New York: The Free Press.
- Srinvasan, V. & Mason, C. H. (1986). Nonlinear least squares estimation of new product diffusion models. Management Science, 5, 169-178.

Biographies



Dr. David R. Firth's research interests focus on understanding innovation in organizations. This interest has lead to several streams of research. One stream is on the use of genre analysis to comprehend organizational communication, and other ways to understand the systems supporting such communication. A second stream, and the focus of his dissertation which was completed in 2003, is on understanding organizational memory systems, in particular how individuals supply, search for and use information within them. The final stream examines the impact IT Research and Analysis firms have on the comprehension and adoption of IT innovations. Dr. Firth has a Ph.D. in Information Systems from the Anderson Graduate School of Management at UCLA, and an MA in Natural Sciences and a BA (honors) in Physics from the

University of Oxford. He also is a Chartered Accountant. Before his Ph.D., he was a senior manager in the IT control and security practice at KPMG LLP. He is presently an Assistant Professor of Information Systems and Technology at The University of Montana



Dr. Cameron Lawrence holds a Ph.D. from the London School of Economics. His research interests are focused on technological mediated organizational transformation, the application of neo-institutional theory into information systems research as well as the organization and management of the IT function within large organizations. His teaching interests are in the areas of systems and operations, network infrastructure management and quantitative methods. While teaching at the London School of Economics he was awarded the highest teaching award that can be bestowed upon Ph.D. candidates for demonstrated excellence in the classroom.



Dr. Shawn F. Clouse is the Information Technology Director and an Assistant Professor in the School of Business Administration at The University of Montana. He earned an MBA from The University of Montana and a doctorate in Higher Education and Technology Leadership from the School of Education at The University of Montana. Shawn's academic affiliation is with the Information Systems and Technology Department where he teaches courses on networking, e-Commerce, and multimedia development. He is responsible for the management of the School of Business Administration's technology support functions, which includes an online learning environment, videoconferencing, multimedia class-rooms, computer labs, and local area networks. His research interests in-

clude distance learning, utilizing technology in the teaching and learning process, and using technology to support e-Commerce and other business processes.