Implicit Affinity Networks

M. Smith[1], C. Giraud-Carrier[1] and B. Judkins[2]
[1]Dept. of Computer Science, Brigham Young University, USA, {cgc@cs.,smitty@}byu.edu
[2]Amazon.com, USA, {brockjudkins@gmail.com}

Abstract
Social networks are typically constructed around an explicit and well-defined relationship among individuals. In this paper, we describe another class of social networks, known as Implicit Affinity Networks (IANs), where links are implicit in the patterns of natural affinities among individuals. Preliminary results with two Web communities, one focused on people's interests and one focused on people's blogs, exhibit rich dynamics and show interesting patterns of community evolution.

1. Introduction
Online communities, also referred to as neo-tribes [3], have sprung up like mushrooms all over the Internet. These communities represent groups of individuals connected by some well-defined, explicit relation, such as a shared medical condition in a health community, a trusted contact link in a business network, or an established friend or family relationship in a photo-sharing community. The resulting social networks are relationship-centered, and their analysis typically assumes that the network is static, or evolves sufficiently slowly to make the study of snapshots relevant and meaningful. Much work has been done to capture, understand, and model the structure of such social networks (e.g., see [13,15]).

Although useful from a practical (computational) standpoint, the assumption of a static network tends to limit the kinds of analyses that may be performed. Recently, some researchers have begun to study the actual dynamics of social network formation and evolution, leading to the discovery of several interesting patterns such as degree power laws and shrinking diameters (e.g., see [4,6,7,11,14]). We propose to go further and allow the nature of the underlying relationship to vary by focusing on implicit affinities. We take an individual-centered rather than a relationship-centered view of social networks. We consider individuals as social actors characterized by a wide range of attributes and we let relationships among them emerge naturally as a result of commonalities across attributes. Unlike traditional social networks where links represent *explicit* relationships, the links in our approach are based strictly on affinities, or inherent similarities, among the social actors, which create *implicit*, and multi-faceted, relationships. We call the resulting networks *Implicit Affinity Networks* (IANs). Because individuals are complex entities whose attitudes and behaviors change over time, IANs are intrinsically dynamic, and evolve naturally with such factors as their participants' age, occupation, interests, and life circumstances.

In this paper, we describe how IANs can be generated from information about individuals, to visualize and analyze affinities among groups of these individuals. We then report on the early evolutionary stages of a Web community based on implicit affinities as well as the richer dynamics of a blog-inspired community.

2. Community Generation: IANs
We represent individuals by collections of attributes and associated value sets. Each attribute captures some information about individuals, such as occupations, hobbies, research interests, birth place, etc. In our context, an individual may be characterized by any number of attributes and each attribute may have any number of its possible values. Whenever two individuals share an attribute whose value sets overlap, we say that there is an *affinity* between them. A group of individuals together with their affinities can be represented as a graph or network, known as an Implicit Affinity Network (IAN), where each node corresponds to an individual and each edge to an affinity. Any time an individual X adds a value, say v, to one of its attributes, say A, new edges are automatically added between X's node and all existing nodes whose individuals have value v for A. Since we are interested in tracking evolution, our networks are

actually time graphs, as defined in [5], where every node and every edge in the network is time-stamped with the time at which it was added.

In principle, any similarity function defined over pairs of individuals may be employed to build an IAN. Our focus, here, is on the analysis of the network rather than the specific underlying similarity function. Hence, we propose a relatively simple function, as follows. Let Γ be a set of attributes, and for each attribute $A \in \Gamma$, let V_A denote the arbitrary value-set of A. For any individual X, let $Attr(X) \subseteq \Gamma$ denote the set of attributes of X, and $V_A(X) \subseteq V_A$ denote the set of values of attribute A for individual X. Then, the affinity score between X and Y is given by:

$$AffScore(X,Y) = \frac{\sum_{A \in Attr(X) \cap Attr(Y)} AffScore_A(X,Y)}{|Attr(X) \cap Attr(Y)|}, \text{ where } AffScore_A(X,Y) = \frac{|V_A(X) \cap V_A(Y)|}{|V_A(X) \cup V_A(Y)|} \times \alpha_A$$

The term α_A is an optional weighting factor for the Jaccard's index $AffScore_A(X,Y)$. This weight may be used to reflect the relative importance of A in a community. In the most general case, α_A is a composite of individual user preferences and a mined community preference. The former is elicited from individuals, e.g., using a kind of 5-star rating. The latter is the ratio of the number of individuals that have at least one value for attribute A to the total number of individuals in the community. It acts as a global, learned, community weight that evolves with changes in the behavior of individuals and favors frequently used attributes.

3. Social Capital for Community Tracking

Several measures have been proposed to capture the structure and evolution of social networks, including nodal degree, diameter and density (e.g., see [15]). Here, we propose a measure, based loosely on the notion of social capital, which originates in political science and sociology (e.g., see [8]). The notion of social capital seems relevant, and rather intuitive, in the context of implicit affinity networks. Social capital fosters reciprocity, coordination, communication, and collaboration. It has been used to explain, for example, how certain individuals obtain more success through using their connections with other people. It has been suggested that "social capital can be viewed as based on social similarity, the shared affiliations or activities that indicate *how* one knows someone" [1]. In this sense, social capital is not limited to explicit relationships but also implicit ones that result from similarities that may exist in individuals' attitudes and behaviors.

Two main components of social capital have been defined: bonding social capital and bridging social capital [9,10]. Bonding social capital refers to the value assigned to social networks among homogeneous groups of people. Bridging social capital refers to the value assigned to social networks among heterogeneous groups of people. Associations and clubs typically create bonding social capital; neighborhoods and choirs tend to create bridging social capital. Whereas bonding social capital increases through closure, as individuals strengthen existing links among themselves, bridging social capital increases through brokerage, as individuals establish new links across structural holes [2].

Because IANs capture implicit affinities, they can only be used to compute the *potential* for social capital rather than social capital itself. Social capital really accrues when individuals are aware of it, that is, when they establish explicit and intentional relationships with each other. It is still informative to understand and track what social capital may be available to individuals and communities.

Here, we define bonding and bridging potentials simply, as reciprocal of each other, by the following formulas, where N denotes the number of nodes and E the set of edges in the network:

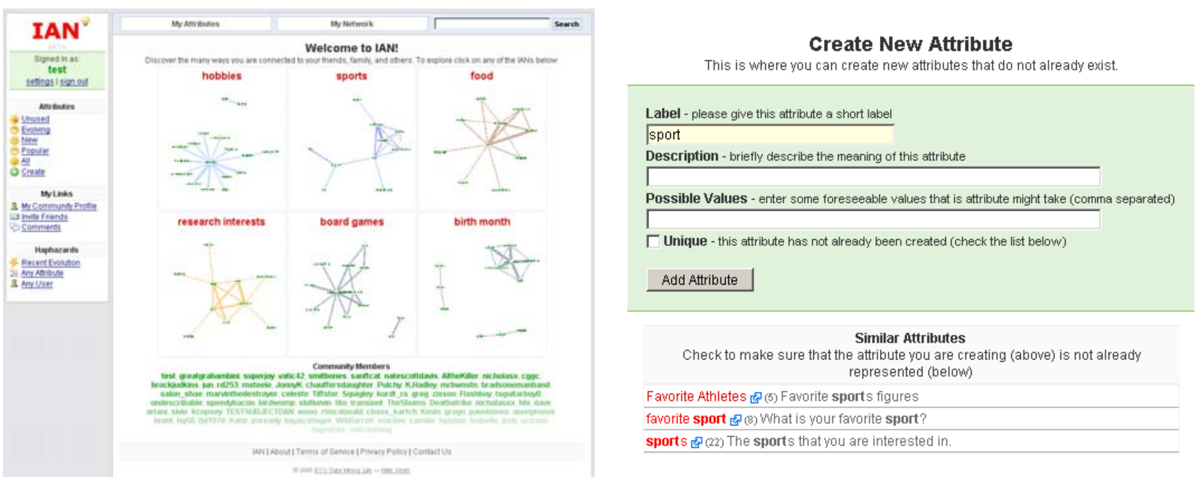
$$BondingPotential = \frac{2}{N(N-1)} \sum_{\{X,Y\} \in E} AffScore(X,Y) \text{ and } BridgingPotential = 1 - BondingPotential$$

Whereas *BondingPotential* is indeed a measure of homogeneity, *BridgingPotential* is a measure of diversity within the network, suggesting how individuals may further connect. Every time an individual adds a new attribute or a new value to an existing attribute, we therefore say that this individual is attempting to bridge out by seeking new connections with new people.

Notice that *BondingPotential* is essentially a weighted version of the density measure Δ defined in [15], where the edges' weights are given by overall affinity scores. One of the unique features of IANs is that these weights are not fixed, but naturally adapted as the profiles of individuals change. Specifically, for an edge $\{X,Y\}$, if X adds a value to one of its attributes, say A_k, and that value is not shared by Y, then the weight of $\{X,Y\}$ decreases.

4. Experiments

As mentioned earlier, most existing online communities are based on a single type of explicit relation among individuals. Even when additional data is available about individuals beyond the relation itself, such data typically lacks the time element necessary to analyze the evolution of implicit affinities required by IANs. Hence, for our first experiment, we implemented a Web application that allows individuals to create and edit their profile in the form of dynamic attribute-value sets, where each attribute captures some characteristic or personal dimension of interest that individuals wish to be represented by and share (see Figure 1). Any and all changes to attribute and attribute-values are time-stamped.



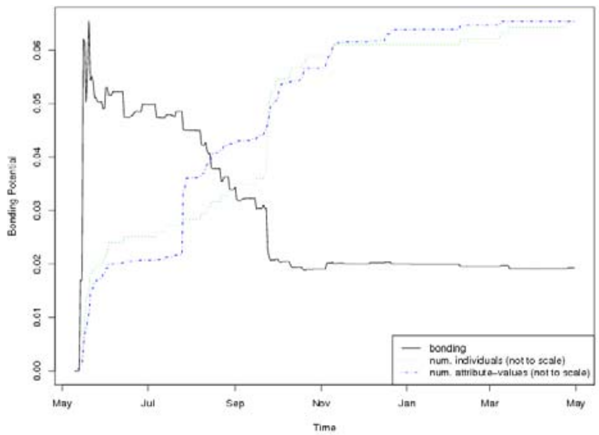
(a) User's Home Page (b) New Attribute Creation

Figure 1: Interests-based Community

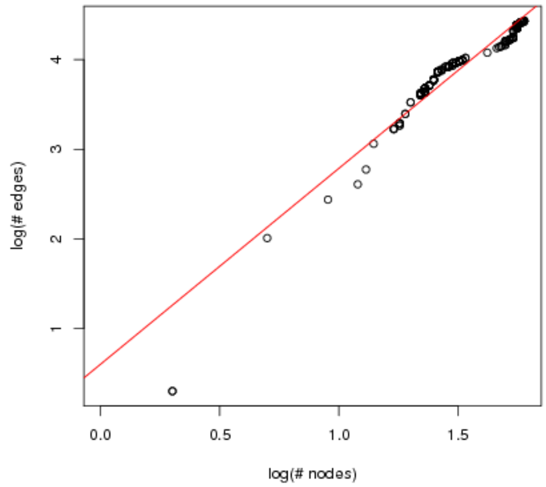
In its current form, the IAN community is a general community that enables sub-communities to emerge over a variety of topics. As of a year from its inception, there were 72 individuals signed up. On average, a user 1) was active within the community for 52 days, 2) visited every 10 days, 3) added 2 attribute-values to their profile per visit, and 4) had 95 attribute-values across 21 attributes.

Recent work on social networks has examined the evolution of the average degree of nodes (i.e., $2|E|/N$) over time [6,7]. We wish to do the same with bonding potential. Figure 2(a) shows the global evolution of bonding potential in the IAN community during our experiment. The number of individuals and the number of attribute-values are also shown to facilitate interpretation. The overall trend in the evolution of bonding potential is decreasing, as might be expected of a still fairly new and rather varied community. At a lower level, the graph may be split into four time periods:

- May. This is the ``birth'' of the community. As one might expect, bonding potential rises as a small number of people join in and begin sharing values on a small set of attributes.
- June-July. This is a period of relative stability, where the number of attribute-values remains relatively constant and only a few new individuals join the community. Note that arrival of new individuals generally results in a short-time drop in bonding potential, followed by an increase, as bonding replaces bridging.



(a) Bonding Potential



(b) #Edges vs. #Nodes

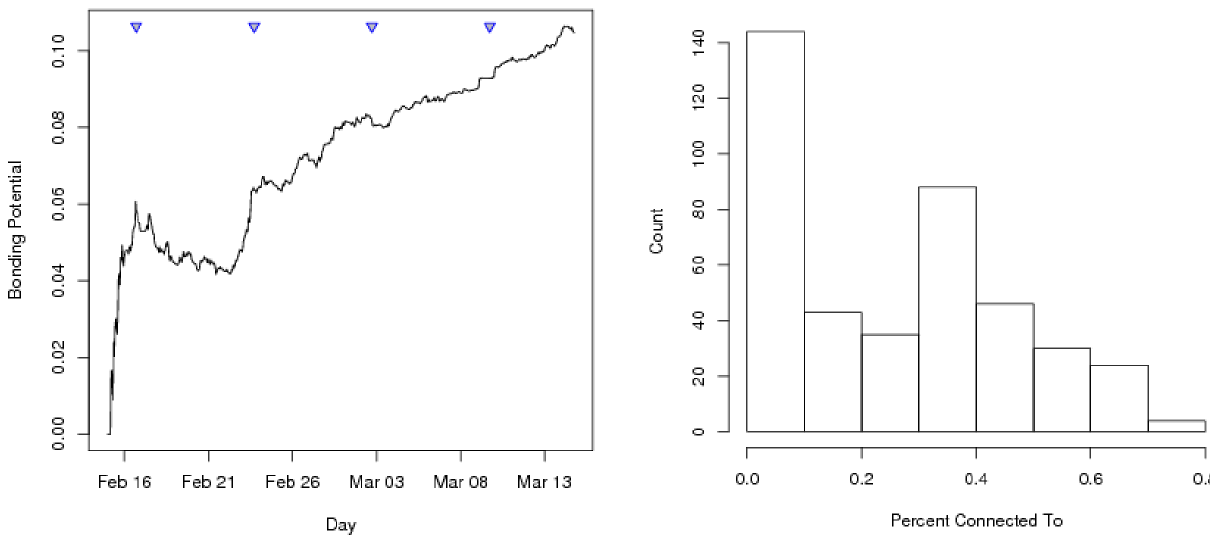
Figure 2: Interests-based Community Evolution

- August-October. This is a period of high activity, partly due to our extending the availability of IAN. During this period a significant number of individuals join IAN and create a significant number of new attribute-values, faster than current members can exploit, thus leading to a decrease in bonding potential. New members are ``casting their lines out," attempting to bridge out by offering new possibilities for affinities with current and new members.
- November-May. As the number of individuals and the number of attribute-values begin to stabilize again, bonding potential plateaus out. The addition of new individuals and new attributes, which causes small troughs on the bonding potential curve, seems to be compensated by the capitalization of individuals on existing attribute-values, i.e., bonding with others rather than bridging out.

Recent studies of dynamic social networks have highlighted characteristics or laws that seem to have broad applicability. In particular, it seems that social networks exhibit densification (i.e., the relation of the number of edges to the number of nodes follows a power law, $E(t)=N(t)^a$ for $1<a<2$) and shrinking diameters (i.e., the 90th percentile of the shortest path lengths between all pairs of nodes decreases over time) [7]. We wish to see whether IANs obey similar laws. We restrict our attention to densification, realizing that our network is still relatively small at this stage. Figure 2(b) plots the log of the number of edges versus the log of the number of nodes, when edges are aggregated across the values of each attribute (i.e., at most one edge per attribute between any 2 nodes). As can be seen, it appears that densification in IANs also follow a power law. Given the way new links arise in IANs, i.e., with probability proportional to the richness of existing individuals' profiles rather than the richness of their connections, we might expect that the exponent be larger than 2. Indeed, the slope of the regression line is 2.18, with R^2=0.92.

In our second experiment, we generate and analyze an implicit affinity network based on blogs. Rather than modeling blog communities based on explicit hyper-linked cross-references as in [5], we model them implicitly, based on blog content. We mine blogs from the public reading list of an influential technology journalist, Robert Scoble [12]. From his list we extracted 19,337 individual blog entries authored by 2,041 bloggers, over a period of a month (from 15 February 2007 to 15 March 2007). To build an IAN from this space of blogs, we represent each blogger as an individual, with a set of attributes and associated values that we mine from the individual's blog entries. Clearly, the more sophisticated the text mining technique used, the richer the description of individuals. For the sake of simplicity, we focus, here, on a single attribute, *Company*, which holds the names of the companies (from a pre-compiled list of 1,914 company names) that may appear in a blog entry. For each blogger, we add the company name value *X* to the attribute *Company* whenever *X* occurs in the body of one of that blogger's entries.

Figure 3(a) shows the global evolution of bonding potential in the IAN community during our experiment. The tick marks at the top of the graph correspond to weekends. The community appears to be bonding overall, as might be expected with a single attribute. Interestingly, however, there appears to be a kind of weekly cyclical activity. It is most visible in the first week, but seems to hold, although to a lesser degree, in the following weeks. The beginning of the week is marked by a decrease in bonding potential, followed by a period of overall increasing bonding, suggesting that bloggers may be bridging out early on and bonding later. This could be due to more prolific or active bloggers that are on the look-out for new information (especially about companies, here) that they add to their blogs early on in the week, followed by a group of less active bloggers (or weekend bloggers) that later catch up with current conversations.



(a) Bonding Potential (b) Connectedness

Figure 3: Blog-based Community Evolution

The number of active bloggers, not shown here, does indeed seem to peak towards the end of the week. Figure 3(b) shows a histogram of the number of bloggers connected to a given fraction of other bloggers in the community. The heavy tail, and the significant number of bloggers connected to between 30% and 40% of the others, suggest that many bloggers are implicitly connected to a relatively high percentage of the community. This is in contrast to the micro-communities found in [5].

One potential benefit of IANs in the blogsphere is that unlike hyper-linked cross-references, IAN links are implicit and therefore may not be known to bloggers. In particular, bloggers may not realize how or where they fit within a particular community based on blog entry content. Thus, an IAN might be used to inform bloggers as to where they reside in the implicit network. For example, are they blogging about things that few others in the community are (i.e., bridging) or are they blogging about the same things that many others are (i.e., bonding opportunity). Knowing where bloggers fit in the IAN today provides information on what they need to do to get where they would like to be in the future.

5. Conclusion

We have shown how to generate a novel class of individual-centered social networks, known as implicit affinity networks. Rather than being built around an explicit relationship, these networks capture dynamic, multi-faceted relationships implicit in the shared characteristics or attributes of individuals. We have discussed the use of the notion of social capital to measure the evolving potential of a community, and have used it to report on experiments with two Web communities, one built around interests and the other around blog content.

In addition to extending the use of IANs to other areas, such as online health communities, to detect trends (e.g., in symptoms, experiences and feelings) that may otherwise remain undetected by physicians, we are working on the idea of overlaying the IAN of a community with the explicit social network (ESN) of the same community. This has already led to the development of measures of bonding and bridging social capital that are not reciprocal. The resulting hybrid network should provide a clean formalism to track the actual social capital of a community more accurately, and thus serve to effectively model important problems in the political and social sciences.

References

[1] Belliveau, M.A., O'Reilly, C.A. III and Wade, J.B. (1996). Social Capital at the Top: Effects of Social Similarity and Status on CEO Compensation, *Academy of Management Journal*, **39**(6):1568-1593.

[2] Burt, R. (2008). Network Duality of Social Capital, in Bartkus, V. and J.H. Davis (Eds.), *Reaching Out, Reaching In: Multidisciplinary Perspectives on Social Capital*, Edward Elgar Publishing.

[3] Johnson, G.J. and Ambrose, P.J. (2006). Neo-Tribes: The Power and Potential of Online Communities in Health Care, *Communications of the ACM*, **49**(1):107-113.

[4] Katz, J.S. (2005). Scale Independent Bibliometric Indicators, *Measurement: Interdisciplinary Research and Perspectives*, **3**:24-28.

[5] Kumar, R., Novak, J., Raghavan, P. and Tomkins, A. (2003). On the Bursty Evolution of Blogspace, in *Proceedings of the 12th International Conference on World Wide Web*, 568-576.

[6] Kumar, R., Novak, J. and Tomkins, A. (2006). Structure and Evolution of Online Social Networks, in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 611-617.

[7] Leskovec, J., Kleinberg, J. and Faloutsos, C. (2005). Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations, in *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 177-187.

[8] Lin, N. (2001). *Social Capital: A Theory of Social Structure and Action*. NY: Cambridge University Press.

[9] Putnam, R.D. (2000). *Bowling Alone: The Collapse and Revival of American Community*, NY: Simon & Schuster.

[10] Putnam, R.D., Feldstein, L.M. and Cohen, D.J. (2003). *Better Together: Restoring the American Community*, NY: Simon & Schuster.

[11] Redner, S. (2005). Citation Statistics from 110 Years of *Physical Review*, *Physics Today*, **58**:49-54.

[12] Scoble, R. (2007). *Scobleizer's Blogs*, online at: http://www.bloglines.com/public/scobleizer.

[13] Scott, J. (2000). *Social Network Analysis: A Handbook*, SAGE Publications.

[14] Tantipathananadh, C., Berger-Wolf, T. and Kempe, D. (2007). A Framework for Community Identification, in *Dynamic Social Networks, in Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 717-726.

[15] Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*, NY: Cambridge University Press.