Group 7

Quantifying Community Growth in Dynamic Social Networks

Denny Britz
dbritz@stanford.edu

Caiyao Mai
cymai@stanford.edu

Chuan Xu
chuanxu@stanford.edu

ABSTRACT

Community detection in social networks is a method that helps us to discover groups of users that are tightly connected. So far, most research has focused on detecting communities in static networks. However, networks evolve over time, and so do the communities within these networks. Understanding the evolution of communities is important to gain insight into macroscopic trends that drive changes in the network. Such an understanding may also help us to predict future growth and trends within the network. In this paper we will present a new model and algorithm for identifying changing, or *dynamic*, communities in evolving networks. We use will use dynamic communities to develop the concept of *community growth rate*, which quantifies how fast a community grows relative to the overall network. We then show that the *community growth rate* is only one example of a family of more general *community evolution metrics*. Finally, we evaluate our model on real-world data using the Angellist dataset, a network of entrepreneurs, and show how the community evolution event we identified make intuitive sense based on our understanding of the domain.

1. INTRODUCTION

Community detection in social and information networks is concerned with finding clusters of users who are densely connected. The identification of communities often reveals interesting properties shared by its members, such as common geographical location, occupation, or social circles. However, the notion that such communities may evolve and change substantially over time has been largely ignored. As networks evolve, we will see that communities may *grow*, *shrink*, *die*, *split*, or *merge*. Studying the evolution of such communities is important for understanding how user interaction patterns evolve, and how user adoption is changing over time. Measuring community growth in dynamic social networks is a relatively new, but very interesting problem. Understanding which communities are growing or shrinking exposes latent trends of user behavior and adoption. Such trends may be a result of external forces, such as competitors or cultural shifts, or internal developments, such as the behavior of existing users within the network. Only by understanding these growth trends is it possible to counteract or accelerate them. Quantifying community growth may also allow us to predict future trends within the network.

This paper is divided into three major parts:

1. *Related Work*. We briefly review preliminaries and present related work in the area of dynamic social network analysis, community detection and community evolution.
2. *Community Growth Rate*. We present a new model to quantify the community growth rate for dynamic communities within an evolving network. The community growth rate describes the growth of a particular community relative to the average network growth at the same time. We present a mathematical definition for the community growth rate and present an algorithm to detect dynamic communities and measure their growth. We then generalize our model to develop a general notion of *network evolution metrics* other than the community growth rate.
3. *Evaluation*. We evaluate our model and algorithm on a real-world dataset of the Angellist network. Angellist is a social network of entrepreneurs, startups and investors that exhibits community structure in geographies and vertical markets. We begin with a brief introduction to the Angellist dataset and present common summary statistics to gain an understanding of the network structure that will help us put our results into perspective. Finally, we evaluate the results of running our algorithm for detecting dynamic communities and quantifying community growth presented in this paper and reason about the validity of our results based on our understanding of the Angellist network.

2. RELATED WORK

A large body of research exists on static network analysis. Power laws, the fact that there exist few nodes with very high degree and a long tail of nodes with a small degree, are common in real world data and have been observed in many empirical studies of social networks [6]. Macroscopic network evolution has also been studied in the past years. It was found that many social networks exhibit densification (average node degree increases) and shrinking diameter properties and that they are composed of three parts, a big component, isolated nodes, and a middle region [5; 7].

A significant amount of work has been done on detecting communities in static graphs. This includes modularity methods [1; 2; 3], spectral clustering methods [4] and more. A comparison of different algorithms for community detection can be found in in [10]. More recently, an approach called BigCLAM for improving the detection of overlapping communities has been proposed [12]. However, all of these methods ignore temporal information and try to detect communities on static graphs.

Recently, more researchers have become interested in mining the temporal evolution of networks, which includes the evolution of communities within these network. The evolution of communities in social networks has been explored from the viewpoint of the individual on a microscopic level [8; 9]. Macroscopic community evolution, which is what we are interested in, has not yet attracted a large body of research. In [15] a framework to identify communities spanning several time steps is presented. Individuals are grouped based on interactions at each time-step, and the problem of matching individuals to communities is solved by using a dynamic programming approach to find colorings for individual nodes. A model for dynamic communities based on existing community detection techniques has been proposed in [11], which defined dynamic communities as an ordered set of at most one community per time-step of the evolving network. We will briefly review the model of dynamic communities in section 3.1 and then extend it with our own developments.

3. COMMUNITY GROWTH RATE

Measuring community growth is challenging for various reasons. For one, static community detection is a difficult problem in itself, and there exist many algorithms that identify different structures of communities. We will see later that one needs to make decisions based on domain knowledge about the network to arrive at sensitive results for dynamic communities. In this section we present a mathematical model for the community growth rate, which quantifies the growth of a particular community relative to the average network growth rate at the same time. Our model is independent of the community detection algorithm, which also means that its accuracy is dependent on the accuracy of the detection algorithm used.

3.1 Review of Dynamic Communities

We briefly review the model of dynamic communities developed in [11]. We will use a generalization of this model in this paper. Intuitively, a dynamic community describes how a community evolves over time. Assume we possess a set of time-step graphs $g_t \in \{g_1, ..., g_k\}$. These are snapshots of the graph at certain time steps, for example, at the beginning of each month. For each time-step we extract a set of step communities $\mathbb{C}_t = \{C_{t1}, .., C_{tm}\}, C_{ti} \subseteq \mathbb{V}$ using any static community detection algorithm. We then define a set of dynamic communities $\mathbb{D}$, where each dynamic community is an ordered set of step communities. Each dynamic community D_i has a front F_i which is the latest step community that D_i was matched to. Note that in this model, a dynamic community can have at most one step community for each time step t.

For our purposes we modify (and generalize) the model and define a dynamic community to be a directed acyclic graph (DAG) of step communities as opposed to an ordered set. See figure 5 for an example of such a graph in our evaluation step. Using this definition, a dynamic community can now contain more than one step community at a time step t, which is not the latest front F. We note that there exist several key events that can happen for a dynamic community at each time step: A new dynamic community can be *born*. An existing dynamic community can *die* when the front of that community has no outward edge. Two two dynamic communities *merge* when they are both parents of the same step community at the next time step. An existing dynamic community can *split* when it is the parent of two or more step communities at the next time step. Finally, an existing existing dynamic community can *expand* or *contract* when it has exactly one child. Both *expansions* and *contractions* are instances of community *continuations*. We present an algorithm for discovering dynamic communities and determining such events in 3.3.

3.2 Defining Community Growth Rate

Equipped with the definition of dynamic communities we are now ready to define the community growth rate. Let D be a dynamic community defined by a DAG of step communities $\mathbb{C}_t = \{C_{t1}, .., C_{tm}\}$. $D = (V, E), V \subseteq \mathbb{C}_t$. Let F_t be the front of D at time t. By definition, the latest front of D consists of exactly one step community. Strictly speaking, the front F is a set of step communities, but for the sake of brevity we will be using F_t as the union of these step communities if contains more than one element. We then define the *absolute community growth rate* of D at a specific time-step t as:

$$r^t(D) = \frac{|F_t|}{|F_{t-1}|}, t > 1 \tag{1}$$

The above tells us how much a community grows or shrinks over time. However, we are more interested in how the community behaves relative to the overall network, and whether it grows grow faster or slower than the network as a whole at the same time step. Therefore, let $\mathbb{V}_i$ be the set of vertices for g_i, the graph at time-step i. We now define the *community growth rate* γ_i at time-step i as

$$\gamma^t(D) = \left(\frac{|\mathbb{V}_t|}{|\mathbb{V}_{t-1}|}\right)^{-1} \frac{|F_t|}{|F_{t-1}|} = \frac{|\mathbb{V}_{t-1}||F_t|}{|\mathbb{V}_t||F_{t-1}|} \tag{2}$$

Intuitively, a community growth rate larger than 1 tells us that the dynamic community D is growing faster than the network at time step i. We also define the *average community growth rate* $\gamma(D)$ for a dynamic community D as:

$$\bar{\gamma}(D) = \frac{1}{t_{max}} \sum_t \gamma^t(D) \tag{3}$$

The average community growth rate is helpful in determining whether a community is growing as a result of a microscopic trend, such as user adoption rapidly increasing in a

1. Apply static community finding algorithms on g_1 to extract $\mathbb{C}_1$. Initialize $\mathbb{D}$ by creating a new dynamic community for each step cluster $C_{1i} \in \mathbb{C}_1$.
2. For each subsequent step $t > 1$, extract $\mathbb{C}_t$ from g_t.
3. Process every $C_{ta} \in \mathbb{C}_t$ as follows:
   (a) Match all dynamic communities D_i for which $sim(C_{ta}, F_i) > \theta$.
   (b) If there are no matches, create new dynamic community containing C_{ta}
   (c) Otherwise, create an edge from each matching dynamic community to C_{ta}
4. Update the set of fronts F for each dynamic community to be the latest matched step community. For each case where one existing dynamic community has been matched to 2 ore more step communities, the front is now the union of these communities.
5. Repeat from #2 until all time step graphs have been processed.

Figure 1: Algorithm for creating a DAG of step communities

certain geographical region. One may use such insights to accelerate or inhibit the behavior.

3.3 The algorithm

Figure 1 presents an algorithm for creating DAGs of dynamic communities. Once we have these DAGs finding community events and the growth rate is trivial. Our algorithm uses a similarity function $sim(C_{ta}, F_i) > \theta$ that compares the front F_i of a dynamic community to a step community from the static network. This raises the question of what makes for a good similarity function. The Jaccard distance of the two sets seems is a natural choice that is recommended in [11]. However, we will later see that the Jaccard distance is not a good choice for rapidly growing networks due the large difference in set sizes from one time step to another. The second parameter of our model is θ, the similarity threshold. Intuitively, a small θ will lead to many *merge* and *split* events due to the high probably of a dynamic community being matched to one or more new step communities. A large θ will lead to more community deaths and continuations. How to determine θ is an open problem and may require domain knowledge of the network. We will describe a way for determining theta experimentally in our evaluation section.

3.4 General Community Evolution Metrics

It is easy to see that the community growth rate $\gamma^t(D)$ is only one element of a family of possible *community evolution metrics*. For example, one may define the change in clustering coefficient or effective diameter from one time step to another in a similar way. More formally, let $\mathbb{M}$ be a set of metrics that act on a network and produce a real-valued number. $\mathbb{M} = \{m_i \mid m_i : (V, E) \to \mathbb{R}\}$. Then we define the set of such *evolution metrics* as Γ, the set of all γ_i such that:

$$\gamma_i^t = \frac{m_i(G_{t-1}) m_i(F_t)}{m_i(G_t) m_i(F_{t-1})} \tag{4}$$

where F_t is the front of the dynamic community D of interest at time t and $G_t = (V_t, E_t)$ is the network at time t as a whole. Similarly, we define the average of such an evolution metric as:

$$\bar{\gamma}_i^t = \frac{1}{t_{max}} \sum_t \gamma_i^t = \frac{1}{t_{max}} \sum_t \frac{m_i(G_{t-1}) m_i(F_t)}{m_i(G_t) m_i(F_{t-1})} \tag{5}$$

As community evolution metrics are possibly correlated, an interesting problem is to explore the effect of a set of metrics on another metric. While we have an understanding of how metrics of static networks are related to each other, this knowledge does not necessarily translate to the community level.

4. EVALUATION ON REAL-WORLD DATA

4.1 Overview

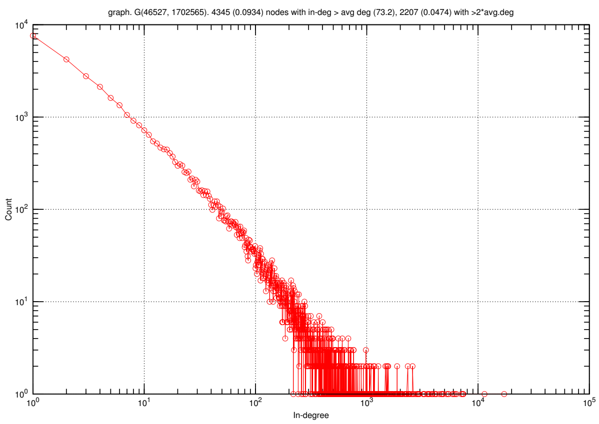
Angellist is a social network of entrepreneurs, startups and investors. Entrepreneurs create profiles for their startups on Angellist, and use the platform to attract followers and investment. Angel investors use the platform to spot investment opportunities and invest directly through Angellist. Angellist was founded in 2010, and has become a common way to raise seed capital. The social aspect of Angellist is similar to that of Twitter. People can follow other people as well as companies. A person who follows another person (or company) is notified about their activity in his or her activity feed. We have decided to use the Angellist dataset for several reasons. Angellist has provided us with edge creation time stamps, which allows us to model the evolution of network. We also have labeled data for geographical and vertical markets, which will help us to interpret our results and compare them to our intuitive understanding of the network. For this purpose we have generated a set of networks $\{g_1, ..., g_k\}$, each representing a monthly snapshot. For our analysis we will be using 12 network snapshots, from January 2011 to December 2011.

4.2 Network summary statistics

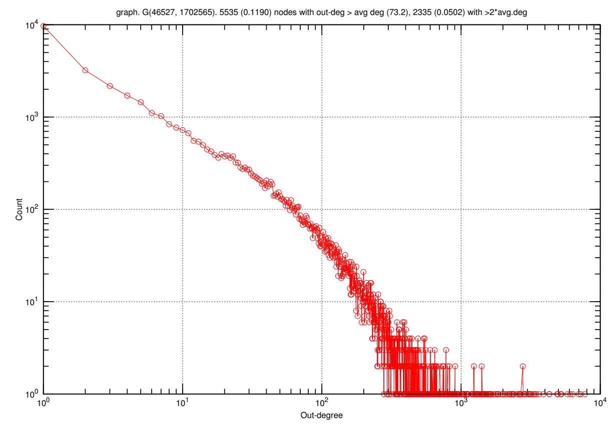
Number of Nodes	46,527
Number of Edges	1,702,565
Number of Undirected Edges	1,407,506
Average Clustering Coefficient	0.235026464643
Effective Diameter	4.51718
Edge Reciprocity	0.34601204653

Table 1: Angellist Descriptive Statistics

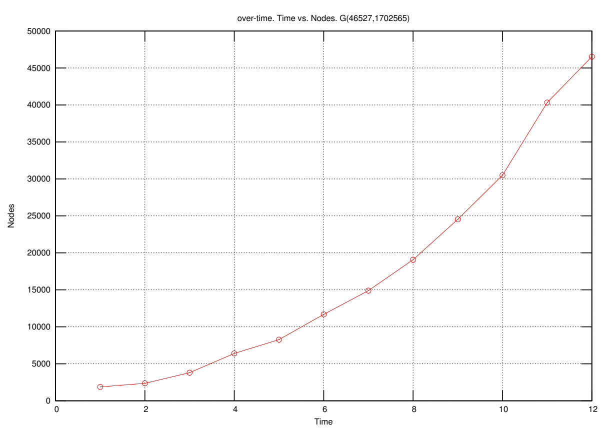
Table 1 shows descriptive statistics of the directed Angellist follower network as of December 2011, the latest step network. The network exhibits a small effective diameter, which is defined as the average number of edges needed to



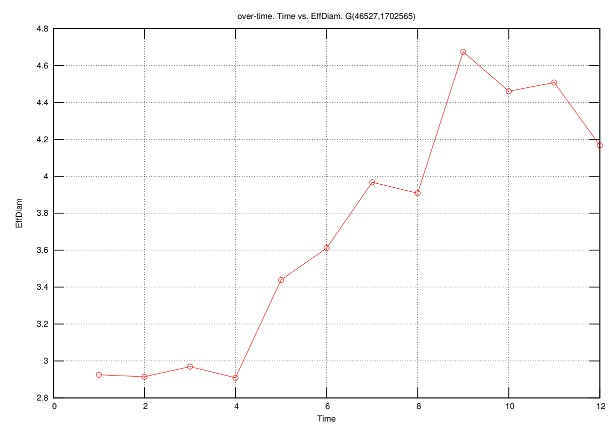
(a) Angellist in-degree distribution



(b) Angellist out-degree distribution



(c) The node arrival process in the Angellist network



(d) The evolution of the effective diameter in the Angellist network

Figure 2: Angellist Summary statistics: In-degree distribution, out-degree distribution, node arrival process, effective diameter evolution.

reach 90% of all other nodes in the network. The diameter is in line with the theory of six degrees of separation, which is common for online social networks. The network has a high clustering coefficient, which again is typical for many online social networks. A high clustering coefficient implies that nodes are tightly knit together, and that people are likely to connect with friends (or followers) of their friends. As a comparison, Orkut exhibits a clustering coefficient of 0.1704 and Flickr a clustering coefficient of 0.3616 [13]. Sometimes it is convenient to treat a directed network as undirected, therefore we calculate the edge reciprocity, which we define as the number of non-unique bi-directed edges, divided by the total number of edges in the graph. It is interesting to see that more than 30% of follows are reciprocated.

As expected, both the in-degree and out-degree distributions of the Angellist network follow a power-law, as can be seen in figures 2a and 2b. This means that a few people have a very large following, but the large long-tail of people have relatively few followers.

Figure 2c shows how new nodes arrive in the network over time. Node arrival is an exponential process, showing fast growth of the user base. The network growth rate will help us to put the community growth rate into perspective. Figure 2d shows the evolution of the effective diameter. We were expecting to find a shrinking diameter as is the case for many social networks, but we find the opposite. One possible explanation is that Angellist started out as a very tight-knit community with an extremely small diameter. As it became more mainstream, communities that are further removed from each other formed, resulting in an increase of diameter. We see that the diameter stays small and will most likely converge to a more stable value at a future point in time.

4.3 Results

We now evaluate our model of community growth on the Angellist dataset. As described above, we divided the Angellist dataset into 12 time-step networks, one snapshot for the beginning of each month ranging from January 2011 to December 2011. We evaluated our model using two different community detection algorithms: Fast Modularity Maximization [14] and BigCLAM (Cluster Affiliation model for Big Networks) [12]). For the BigCLAM algorithm we used cross-validation techniques to determine the most likely number of communities. For both algorithms, we only considered communities that have a significant size of more than 100 nodes. Figures 3a and 3b show the number of communities and the the sizes of these communities as detected by each of the algorithms over time, respectively. We can see that BigCLAM identified more communities than the modularity maximization algorithm. However, since BigCLAM focuses on identifying overlapping communities, these are not necessarily distinct from each other and two communities may have share a large number of users.

Next, we implemented the algorithm presented in section 3.3 for both community detection algorithms. Figure 2c shows that the Angellist network is growing fast. As a result of that, we found that the Jaccard distance was not a good choice for the similarity metric. For example, if the networks doubles in size and we assume that communities

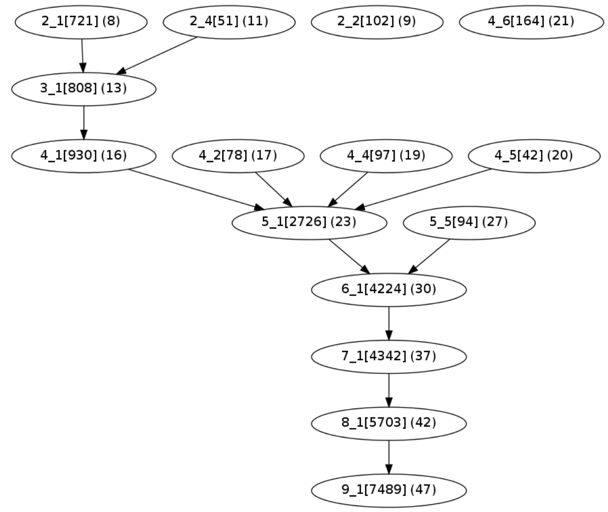


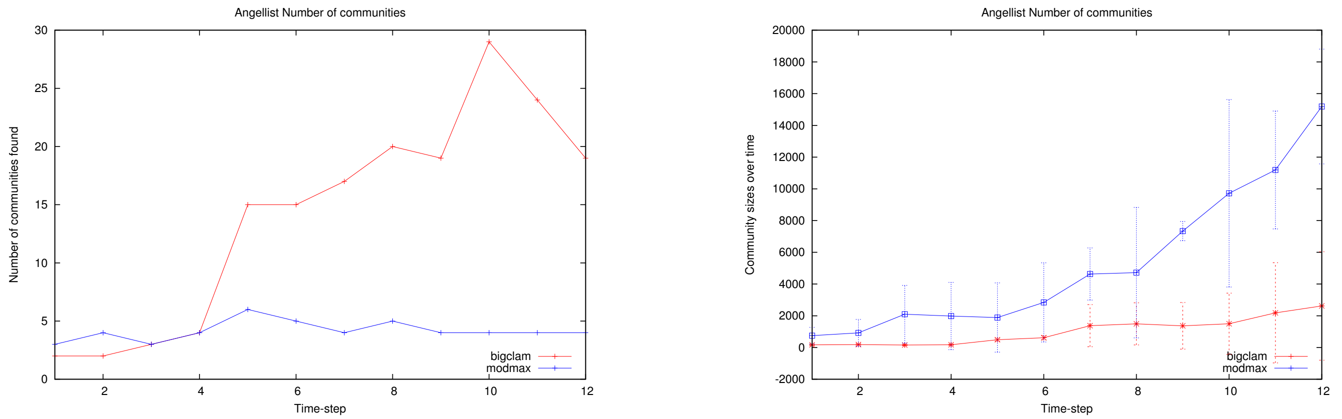
Figure 5: A growing dynamic community identified by modularity maximization. The labels are of format [time]_[community_id] [size] (node_id). The community continues to grow after time step 9 (not shown)

grow at about the same rate, then the Jaccard distance will be at most 0.5 due to the difference in set cardinalities alone. Instead, we measured how many of the nodes in community c_1 are still together in a community c_2 at the next time step:

$$sim(c_1, c_2) = \frac{|c_1 \cap c_2|}{|c_1|} \quad (6)$$

Such a metric makes intuitive sense for a rapidly growing network. If a community c_1 is a subset of a community c_2 at the next time step, then it is likely that newly arriving nodes in c_2 have joined the community c_1. For a network that is not rapidly growing the Jaccard distance may be a better choice as a similarity metric.

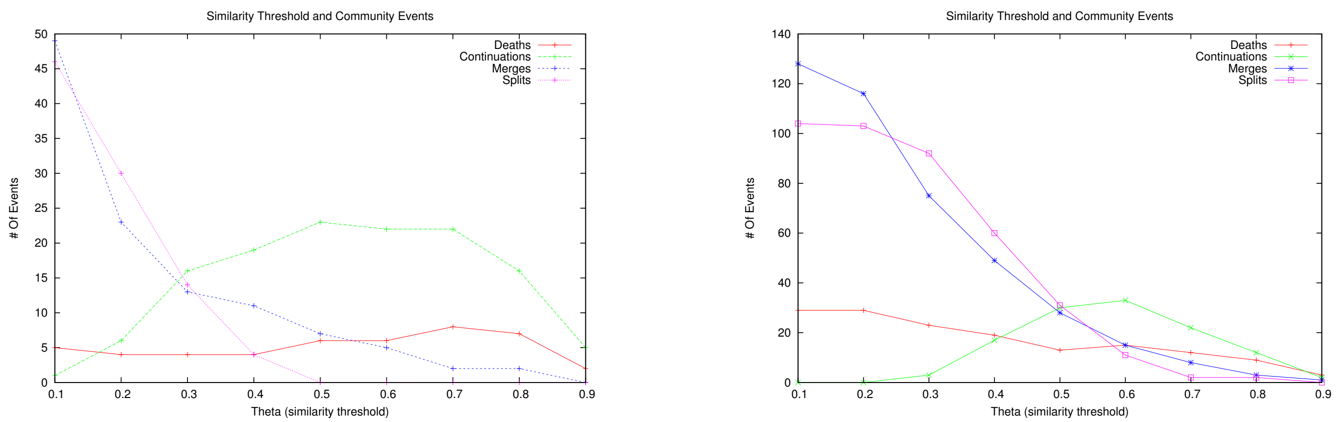
We also experimented with different values of the similarity threshold θ. Figure 4 shows the number of dynamic community *splits*, *merges*, *continuations* and *deaths* as a function of θ for both algorithms. Based on an intuitive understanding of the network evolution we believe that continuations should be more frequent than splits or merges. Therefore, we decided that t $\theta = 0.6$ is a reasonable choice for our network. However, this is not a general result. Depending on the domain of the network, one may argue that merges, splits or deaths should be the most common events. The overlap of communities plays another important role in picking θ. If one expects much overlap in the community structure then choosing a lower θ is may be more reasonable as that will preserve distinct communities as opposed to merging them. If the network exhibits little overlap in community structure then a higher θ may yield better results due to the many distinct users in the communities. We believe that there is no general rule for choosing an optimal θ. Instead, one should employ domain knowledge about the semantics of the network and community structure to pick a suitable value for one's use case.



(a) Number of communities detected by modularity maximization and BigCLAM, over time.

(b) Average community size and standard deviation detected by modularity maximization and BigCLAM, over time.

Figure 3: Count and sizes of detected communities by by modularity maximization and BigCLAM



(a) Similarity threshold θ and community events for Modularity Maximization

(b) Similarity threshold θ and community events for BigCLAM

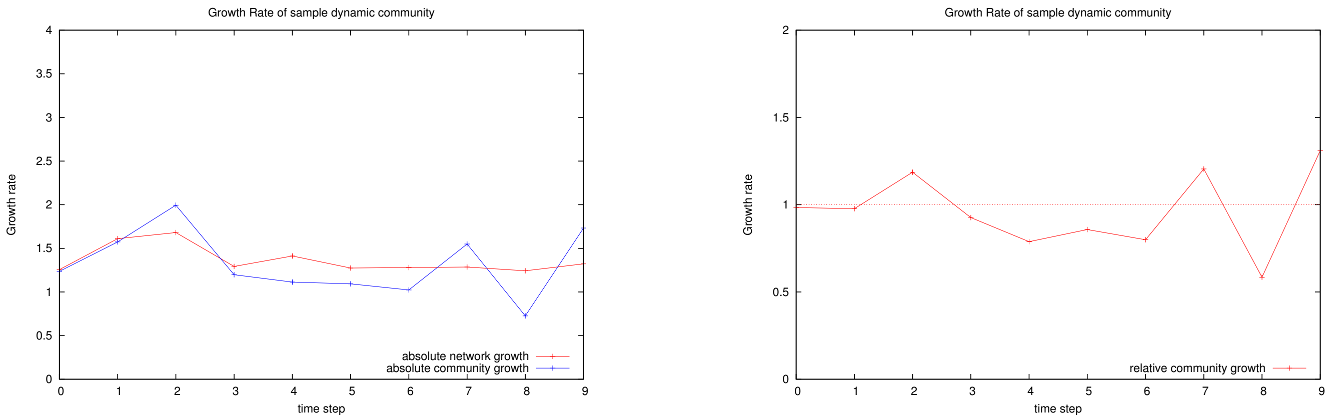
Figure 4: Similarity threshold θ and community events

Figure 5 shows a sample dynamic community that was identified by our algorithm. The community shown grows from 721 nodes at time step 2 to 7489 nodes at time step 9. The community continues to grow after time step 9 without being involved in any merge or split events. This is not shown in the picture for the sake or brevity. We can see that at time steps $t = 4$ and $t = 5$ several communities merge into one large community, which then continues to grow over time.

Figure 6a shows the absolute growth rate for a sample dynamic community and the network as a whole, and 6b shows the relative community growth rate γ^t for the same community. The average growth rate of the shown community is 0.95, which means that the average growth is very close to the total network. However, We see that it varies over time and does not closely follow the overall trend in the network. We have not investigated why this particular instance of a dynamic community behaves the way it does, but it likely is a result of external trends in user adoption. For example, if the owner of the network launched a marketing campaign that encouraged a specific segment of users to join the network, then unrelated communities will grow slower than communities that are joined by such users. Similarly, if the owners of the network advertised to user similar to those in the shown community, then it will likely grow faster than the network as a whole.

In the Angellist network we observed that the communities were initially separated based on geographies. Large geographic communities were dominated by one geographical group. For example, in January, 50% of the largest community were users located in Silicon Valley, and 64% of the second largest community were users located on the East cost. This pattern started to break down in June, when these communities started to merge. The new communities were no longer dominated by one geographic region: Both Silicon Valley users and East Coast users make up about less than 30% of the new mixed community. Based on these findings we conclude that geographical homogeneity of communities gradually increases as a function of time.

Another observation is that communities dominated by one geographical group tend to grow faster than others. We



(a) Absolute network and community growth of a sample dynamic community

(b) Community growth rate γ over time for a sample dynamic community

Figure 6: Sample community growth

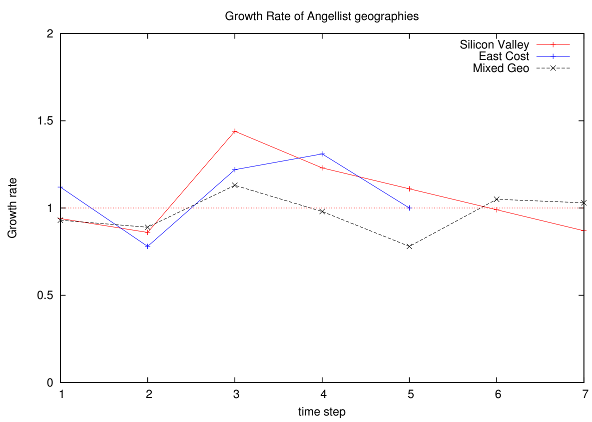


Figure 7: Growth rate of geographically-dominated dynamic communities. At $t = 6$ the East Coast community merges into a mixed community, and no East-Coast dominated dynamic community is left in the network.

evaluated the growth of three sample dynamic communities: One dominated by Silicon Valley users, one dominated by East Cost users, and a third not dominated by a single geographical group. We defined a community as dominated by a geographical region when 50% or more of its users are located in one region. Figure 7 shows the growth rates of these communities over time. Based on these observations we hypothesize that network effects due to geographic co-location are significant in accelerating community growth.

5. CONCLUSION

In this paper, we have developed a model for the community growth rate, which quantifies growth of a particular community relative to the average network growth at the same time. We based our model on the notion of dynamic communities [11] which we generalized from ordered sets to sets of directed acyclic graphs (DAGs) of evolving communities. We used these DAGs to define community evolution events such as *continuation*, *merging* and *splitting*, and *death*. We then presented an algorithm to build DAGs of such dynamic communities and find their corresponding evolutionary events. Finally, we proposed general notion of *community evolution metrics*, of which community growth rate is only one example. Such community evolution are useful in understanding various dimensions of the community evolution structure. We evaluated our model and algorithm on a real-world dataset from Angellist, a network of entrepreneurs, that we divided into 12 time-step networks. We successfully identified dynamic communities, measured evolutionary events and used them to pick a suitable similarity threshold θ. We inspected and reasoned about the evolution of an interesting sample community. Our findings lead us to believe that the evolutionary analysis of dynamic communities can play an important role in helping network owners encourage user behavior they would like to see in the network.

6. FUTURE WORK

We believe there exist many potential avenues for future work to extends our developments and will propose several different directions this section.

6.1 Predictive Models

We believe that it is possible to predict community events as well as the *community growth rate* by training a machine learning classifier on community-level features. Examples of such community level features include clustering coefficient, fraction of out-links, degree distribution, diameter, as well other *community evolution metrics* from previous time steps. Making such predictions could play an important role in accelerating or inhibiting community growth and encourage the desired behavior for the owners of the network.

6.2 Other evolution metrics

While we presented a general notion of *community evolution metrics* we did not present any particular metrics other than the *community growth rate*. We believe that many other interesting metrics can be discovered and evaluated in a similar fashion. Examples of such metrics include clustering coefficient evolution, diameter evolution, degree distribution evolution, or a notion of how isolated community are from the rest of the network, similar to the notion of network expansion.

6.3 Picking a similarity function and θ

We picked the both the similarity function and similarity threshold θ based on our intuitive understanding of the network. We believe that one should be able to build an ontology of similarity functions and values of θ suitable for different kinds of network. It may also be possible to automatically pick a suitable similarity function and threshold by classifying the structure of the network.

6.4 Comparing detection algorithms

We evaluated out model using two community detection algorithms, fast modularity maximization and BigCLAM. Many other community detection algorithms exist, and it would be interesting to investigate how the choice of detection algorithms influences the dynamic communities found in the network. It may also be possible to combine several algorithms to match dynamic communities with a higher accuracy.

7. REFERENCES

[1] Newman, Mark EJ. "Modularity and community structure in networks." Proceedings of the National Academy of Sciences 103, no. 23 (2006): 8577-8582.

[2] Clauset, Aaron, Mark EJ Newman, and Cristopher Moore. "Finding community structure in very large networks." Physical review E 70, no. 6 (2004): 066111.

[3] Chen, Jiyang, Osmar R. Zaane, and Randy Goebel. "Detecting Communities in Social Networks Using Max-Min Modularity." SDM. Vol. 3. No. 1. 2009.

[4] Smyth, S., and S. White. "A spectral clustering approach to finding communities in graphs." In Proceedings of the 5th SIAM International Conference on Data Mining, pp. 76-84. 2005.

[5] Kumar, Ravi, Jasmine Novak, and Andrew Tomkins. "Structure and evolution of online social networks." In Link Mining: Models, Algorithms, and Applications, pp. 337-357. Springer New York, 2010.

[6] Clauset, Aaron, Cosma Rohilla Shalizi, and Mark EJ Newman. "Power-law distributions in empirical data." SIAM review 51, no. 4 (2009): 661-703.

[7] Leskovec, Jure, Jon Kleinberg, and Christos Faloutsos. "Graph evolution: Densification and shrinking diameters." ACM Transactions on Knowledge Discovery from Data (TKDD) 1, no. 1 (2007): 2.

[8] Backstrom, Lars, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. "Group formation in large social networks: membership, growth, and evolution." In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 44-54. ACM, 2006.

[9] Leskovec, Jure, Lars Backstrom, Ravi Kumar, and Andrew Tomkins. "Microscopic evolution of social networks." In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 462-470. ACM, 2008.

[10] Lancichinetti, Andrea, and Santo Fortunato. "Community detection algorithms: a comparative analysis." Physical review E 80, no. 5 (2009): 056117.

[11] Greene, Derek, Dnal Doyle, and Pdraig Cunningham. "Tracking the evolution of communities in dynamic social networks." In Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on, pp. 176-183. IEEE, 2010.

[12] Yang, Jaewon, and Jure Leskovec. "Overlapping community detection at scale: A nonnegative matrix factorization approach." In Proceedings of the sixth ACM international conference on Web search and data mining, pp. 587-596. ACM, 2013.

[13] Hardiman, Stephen J., and Liran Katzir. "Estimating clustering coefficients and size of social networks via random walk." In Proceedings of the 22nd international conference on World Wide Web, pp. 539-550. International World Wide Web Conferences Steering Committee, 2013.

[14] Clauset, Aaron, Mark EJ Newman, and Cristopher Moore. "Finding community structure in very large networks." Physical review E 70, no. 6 (2004): 066111.

[15] Tantipathananandh, Chayant, Tanya Berger-Wolf, and David Kempe. "A framework for community identification in dynamic social networks." In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 717-726. ACM, 2007.