What Can the Temporal Social Behavior Tell Us? An Estimation of Vertex-Betweenness Using Dynamic Social Information

Jing-Kai Lou
Department of EE
National Taiwan University
Email: kaeaura@iis.sinica.edu.tw

Shou-de Lin
Department of CSIE
National Taiwan University
Email: sdlin@csie.ntu.edu.tw

Kuan-Ta Chen
Institute of Information Science
Academia Sinica
Email: ktchen@iis.sinica.edu.tw

Chin-Laung Lei
Department of EE
National Taiwan University
Email: lei@cc.ee.ntu.edu.tw

***Abstract*—The vertex-betweenness centrality index is an essential measurement for analyzing social networks, but the computation time is excessive. At present, the fastest algorithm, proposed by Brandes in 2001, requires $\mathcal{O}(|V||E|)$ time, which is computationally intractable for real-world social networks that usually contain millions of nodes and edges.**

In this paper, we propose a fast and accurate algorithm for estimating vertex-betweenness centrality values for social networks. It only requires $\mathcal{O}(b^2|V|)$ time, where b is the average degree in the network. Significantly, we demonstrate that the local dynamic information about the vertices is highly relevant to the global betweenness values. The experiment results show that the vertex-betweenness values estimated by the proposed model are close to the real values and their rank is fairly accurate.

Furthermore, using data from online role-playing games, we present a new type of dynamic social network constructed from in-game chatting activity. Besides using such online game networks to evaluate our betweenness estimation model, we report several interesting findings derived from conducting static and dynamic social network analysis on game networks.

***Index Terms*—Betweenness, MMORPG, Text-Conversation**

I. Introduction

Betweenness centrality [1], one of the most important measurements for analyzing social networks, ranks vertices according to their global positions in the network. More specifically, it is used to interpret the relative importance of vertices, which serve as bridges connecting different parts of the network.

However, calculating the betweenness centrality for all the vertices in a network is time consuming because it involves finding shortest paths between all pairs, which takes $\Theta(|V|^3)$ time with the Floyd-Warshall algorithm. To improve the time complexity, Brandes [2] proposed a BFS-based search algorithm that requires $\mathcal{O}(|V||E|)$ time, where $|V|$ is the number of vertices and $|E|$ is the number of edges in a network. However, at present, most well-known online social networks, such as MySpace, Flickr, Twitter and other large-scale online communities, consist of millions of vertices and edges. Thus, even with the Brandes algorithm, computing the vertex-betweenness centrality scores for such large-scale networks is still computationally intractable.

This work tries to resolve the above issue. First of all, we consider that a social network is more than a cumulative summary of the interactions among actors in a certain period. Instead, we emphasize the dynamics of a network, such as the time when vertices joined and the links were generated. Networks that carry dynamic social information are called *dynamic social networks* [3]. To describe the evolution of each vertex in a dynamic social network, we propose several local dynamic indicators. Along with a fast and accurate method for estimating the value of vertex-betweenness centrality. The time complexity of our approach is $\mathcal{O}(b^2|V|)$, where b is a small number equivalent to the average degree of the nodes in the network. For most large-scale social networks, b is far smaller than the number of edges $|E|$. For example, for the networks described in [4], the average number of edges is $97,719,006$, but the average degree is only 14.79.

To collect useful dynamic information about the vertices and their neighbors, we introduce two local indicators: *attraction* and *dynamic transitivity*. *Attraction* models the capability of a vertex to attract connections from new vertices; and *dynamic transitivity* measures an entity's capability as a mediator to refer one of its friends to another. We then use a linear regression model to integrate the two variables to predict the global vertex-betweenness values in a dynamic social network.

We demonstrate the efficacy of the proposed fast estimation approach in a practical application by identifying the influential individuals in game social networks. According to the MMO-Chart website [5], there has been a dramatic increase in the number of online-game subscriptions worldwide during the last 10 years. Although more players are joining massively multiplayer online games (MMOGs) and establishing virtual social communities, our approach attaches more importance to identifying influential individuals.

Our contribution in this work is two-fold:

1) We report an interesting phenomenon whereby it is possible to use certain local dynamic indicators to predict global betweenness centrality values. Our experiments show that the values estimated by the proposed model are close to the real values, and their rank is fairly accurate.
2) We propose a new type of dynamic social network, called a MMORPG-network, which is constructed from in-game chatting activity. Besides using the network to evaluate our method for estimating vertex-centrality, we also report some static and dynamic analysis results for it.

II. Dynamic Social Networks

A social network is comprised of a set of vertices joined by edges. Traditionally, such networks summarize the social interactions or relations between the actors within a given period in a static manner. They provide a cumulative rather than temporal overview of the interactions and relations among entities. Hence, it is difficult to answer questions like "Is the network growing or shrinking?" and "Is the speed of change fast or slow?" based on only static information about social networks.

To capture dynamic information, it has been proposed that each vertex and edge should be time-stamped the first time it appears in a social network. As mentioned earlier, networks that provide such temporal information are called *dynamic social networks*. In this work, a dynamic social network is defined as a directed graph $G(V, E)$, where V is a set of vertices and E is a set of edges. The time-stamp of a link from vertex i to vertex j is denoted as τ_{ij}. We use τ_i to denote the time-stamp of a vertex i, i.e. the time it joined the network.

A pair of vertices i and j are said to be *bi-directionally linked* and denoted as $i \leftrightarrow j$ if and only if there exist two opposite-directional edges between i and j. We use $\mathcal{C}_{ij}$ to denote the *channel-establishment time* for a pair of bi-directionally linked vertices i and j. That is, $\mathcal{C}_{ij}$ represents the time the bidirectional edge was established:

$$\mathcal{C}_{ij} = \max_{i \leftrightarrow j}\{\tau_{ij}, \tau_{ji}\}$$

We use $\mathcal{BN}(i)$ to denote the bidirectional-linked neighborhood of vertex i. $\mathcal{BN}$ is defined as follows:

$$\mathcal{BN}(i) = \{j \in V(G) | i \leftrightarrow j\}$$

III. A Fast Algorithm For Estimating Vertex-Betweenness

A. Previous Work on Vertex-Betweenness

Vertex-betweenness is a kind of centrality index that ranks the vertices in a network according to their connectivity. It is usually regarded as a measurement to quantify the importance of a vertex in terms of it connections to other vertices in the network. Following the formal definition proposed by Freeman [1], let p_{ij} denote the number of shortest paths from vertex i to vertex j, and let p_{ijk} denote the number of shortest paths from vertex i to vertex j that pass through the intermediary vertex k. The betweenness centrality score is normalized by dividing the number of pairs of vertices that do not include k as follows:

$$bt(k) = \frac{1}{(n-1)(n-2)} \sum_i \sum_j \frac{p_{ijk}}{p_{ij}},$$

where n is the number of vertices in the network.

B. Limitations

Since calculating the vertex-betweenness values of all vertices in a graph involves finding all-pairs of shortest paths, the operation generally takes $\Theta(|V|^3)$ time with the Floyd-Warshall algorithm. On a sparse graph, it is more efficient to use Johnson's algorithm, which takes $\mathcal{O}(|V|^2 \log |V| + |V||E|)$ time. For a unweighted graph, Brandes [2] proposed an algorithm that requires $\mathcal{O}(|V||E|)$. However, it is still computationally intractable for real-world social networks, which usually consist of millions of nodes and edges.

C. Using local dynamic features to estimate vertex-betweenness

We proposed a fast algorithm for estimating vertex-betweenness. Intuitively, the evolution of a vertex can affect its connectivity in a dynamic social network. We believe it is reasonable to consider the dynamics between a vertex and its neighbors as clues to infer the vertex's global vertex-betweenness value. Therefore, we design a regression model based on the proposed local descriptors to predict such values.

Basically, the local descriptor of a vertex tires to collect local dynamic information about the vertex. Our intuition is that vertices that can attract new comers in every time period have greater potential to achieve high-betweeness scores. Moreover, a vertex that plays the role of local mediator is more likely to possess high betweenness values. Therefore, we propose two local descriptors: *attraction* and *dynamic transitivity*. *Attraction* captures a vertex's ability to attract new vertices to join the network; and *dynamic transitivity* indicates the willingness of a node to introduce their acquaintances to each other. Formally, local descriptors are defined as follows:

Definition 1 (Attraction). *Attraction shows how capable a vertex is in terms of attracting fresh members, and the fresh member is defined as the vertex joining the networks for no more than i days. The indicator of v is the number of bi-directionally linked neighbors who are fresh. We denote the attraction to i-fresh vertices of v as $\mathcal{N}_i(v)$, and calculate it by*

$$\mathcal{N}_i(v) = |\{u \in \mathcal{BN}(v) | \mathcal{C}_{uv} \leq \tau_u + i\}|.$$

If i is set to infinite, the attraction value would be equivalent to the degree value. Although the degree value captures the connectivity of a vertex, it does not indicate how well a vertex can attract new comers.

Definition 2 (Dynamic Transitivity). *Given three vertices v, u, and w, linked bi-directionally in a pairwise manner, if $\mathcal{C}_{vu}$ and $\mathcal{C}_{vw}$ both appear earlier than $\mathcal{C}_{uw}$, then pair (u, w) is regarded as being introduced by vertex v. The dynamic transitivity of v, $\mathcal{T}(v)$, is defined as the number of pairs introduced by it normalized by its clustering coefficient [6] denoted as cc*

$$\mathcal{T}(v) = \frac{|\{(x \in \mathcal{BN}(v), y \in \mathcal{BN}(v)) | \mathcal{C}_{xy} \geq \max\{\mathcal{C}_{vx}, \mathcal{C}_{vy}\}\}|}{cc(v)},$$

where $cc(v) = \frac{2|\{e_{uw}|e_{uv},\, e_{wv},\, e_{uw} \in E(G)\}|}{deg(v) \cdot (deg(v)-1)}$. The term $deg(v)$ equals to the cardinality of $\{x | x \rightarrow v \vee x \leftarrow v\}$.

Note the major difference between the numerator and denominator is that the former considers the order of appearance or edges, but the latter does not, as shown by the example in Figure 1. The channel-establishment time between u and w, $\mathcal{C}_{uw} = 6$, which occurs later than $\mathcal{C}_{vu} = 3$ and $\mathcal{C}_{vw} = 1$. Moreover, $\mathcal{C}_{mw} = 5$ occurs after $\mathcal{C}_{vm} = 3$ and $\mathcal{C}_{vw} = 1$; and the clustering coefficient of v is $2/3$. Thus, the dynamic transitivity of v, $\mathcal{T}(v)$ is equivalent to 3, since the two existing pairs, (u, w) and (m, w), are regarded as being introduced by v.

Note that the usage of cc as the normalization factor does have its empirical justification. From an experiment we conducted, we find a negative correlation between the vertex-betweenness scores and the clustering coefficients of the vertices, given similar out-degrees. Figures 2 shows that the correlation declines from -0.01 to -0.84 as the vertices

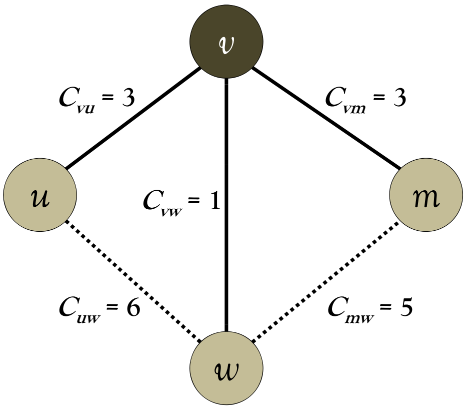


Fig. 1. $\mathcal{C}_{uw} = 6$ is later than both $\mathcal{C}_{vu} = 3$ and $\mathcal{C}_{vw} = 1$. That is, the (u, w) pair is inferentially introduced by v. In addition, $\mathcal{C}_{mw}$ is later than $\mathcal{C}_{vm}$ and $\mathcal{C}_{vw}$. The clustering coefficient of v is $\frac{2}{3}$. Hence, the dynamic transitivity of v, $\mathcal{T}(v)$ is equivalent to 3

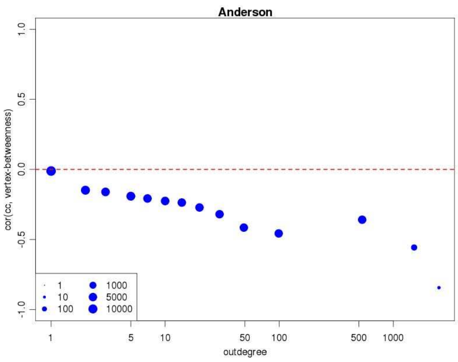
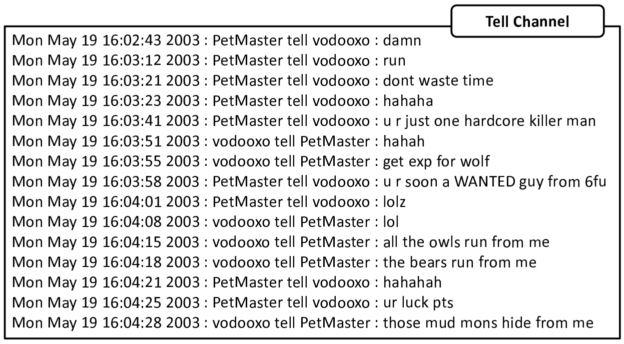


Fig. 2. Correlation between the clustering coefficients and the vertex-betweenness values for given out-degrees. The size of each point represents the number of vertices with a given out-degree



Tell Channel

Mon May 19 16:02:43 2003 : PetMaster tell vodooxo : damn
Mon May 19 16:03:12 2003 : PetMaster tell vodooxo : run
Mon May 19 16:03:21 2003 : PetMaster tell vodooxo : dont waste time
Mon May 19 16:03:23 2003 : PetMaster tell vodooxo : hahaha
Mon May 19 16:03:41 2003 : PetMaster tell vodooxo : u r just one hardcore killer man
Mon May 19 16:03:51 2003 : vodooxo tell PetMaster : hahah
Mon May 19 16:03:55 2003 : vodooxo tell PetMaster : get exp for wolf
Mon May 19 16:03:58 2003 : PetMaster tell vodooxo : u r soon a WANTED guy from 6fu
Mon May 19 16:04:01 2003 : PetMaster tell vodooxo : lolz
Mon May 19 16:04:08 2003 : vodooxo tell PetMaster : lol
Mon May 19 16:04:15 2003 : vodooxo tell PetMaster : all the owls run from me
Mon May 19 16:04:18 2003 : vodooxo tell PetMaster : the bears run from me
Mon May 19 16:04:21 2003 : PetMaster tell vodooxo : hahahah
Mon May 19 16:04:25 2003 : PetMaster tell vodooxo : ur luck pts
Mon May 19 16:04:28 2003 : vodooxo tell PetMaster : those mud mons hide from me

Fig. 3. A snapshot of the Tell-channel log. The log provides the time-stamp of the interaction, the speaker's ID, the listener's ID, and the message's content.

out-degrees increase. In other words, the vertices with the high vertex-betweenness usually have low clustering coefficients, especially to high degree vertices. While the transitivity score of a vertex describes the density in its neighborhood, which are promoted by the vertex. The dynamic transitivity score indicate the ratio of successfully introduced triangles over all triangles around the neighborhood.

D. Regression Model for estimating Vertex-betweenness scores

To estimate vertex-betweenness scores, we build a multiple linear regression model based on the proposed local descriptors. We only exploit two features in the model: attraction and normalized dynamic transitivity.

$$bt(v) = w_1 \cdot \mathcal{N}_i(v) + w_2 \cdot \mathcal{T}(v).$$

The primary time-consuming operation of our scheme is to calculate the dynamic descriptors: attraction and dynamic transitivity. To compute the attraction of all vertices, it requires $\mathcal{O}(|V|)$ time. As for dynamic transitivity, although the approximating algorithm proposed by Schank et al. [7] produce the clustering coefficient in $\mathcal{O}(|V|)$ time, finding the pairs of introduced neighbors for a given vertex still takes $\mathcal{O}(|b^2|)$ time. Therefore, it takes $\mathcal{O}(b^2|V|)$ time to compute dynamic transitivity. As a result, our model requires $\mathcal{O}(b^2|V|)$ time to predict vertex-betweenness values.

IV. Experiment Results

In this section, we describe an experiment conducted to evaluate our approach using a dynamic social network constructed from Fairyland Online game data. The evaluation matrices exploited are the coefficient of determination, Spearman's coefficient, and Kendall's coefficient.

A. Dataset Description

Fairyland Online, which was developed by Lager Network Technologies, operates in Taiwan, Hong Kong, Mainland China, Thailand, and South Korea. The game was launched in Taiwan in February 2003, and attracted more than 200,000 subscribers in less than two months. We collected the chat logs for all realms of Fairyland Online starting from February 2003 to April 2004. The game is comprised of eight independent realms called Alice, Anderson, Candy, Doll, Green, Mermaid, Red, and Wolf.

1) Chat Activity Network: As shown in Figure 3, a *private message* in Fairyland Online is comprised of three tuples in one record: the time stamp of the interaction, the speaker's ID, and the listener's ID. We construct a dynamic social network based on the tuples. A vertex represents an individual and a directed edge denotes a chat message initiated by the source node to the target. The time stamp of the first message from individual i to individual j is preserved as the time stamp of the directed edges (i, j). We call this network a *chat activity network*. However, since the jointime of a vertex i is not given in the data, we deduce the jointime, τ_i, according to the earliest presence of its links. In the other words,

$$\tau_i = \min_{j \neq i, (i,j) \vee (j,i) \in E(G)} \{\tau_{ij}, \tau_{ji}\}.$$

B. Baseline Scheme

For comparison, we propose a baseline model using four common features: out-degrees, in-degrees, bi-degrees, and clustering coefficients to estimate the betweeness centrality. An out-degree captures the activeness of an entity; an in-degree indicates the entity's popularity to some extent; a bi-degree represents how often a two-way communication channel exists; and the clustering coefficient describe the difficulty to be a communication bridge.

We exploit a linear regression model to incorporate these four variables into our prediction model for betweeness centrality:

$$bt(v) = w_1 \cdot od(v) + w_2 \cdot id(v) + w_3 \cdot bd(v) + w_4 \cdot cc(v)$$

C. Evaluation

To roughly see the goodness-of-fit, we compare the true betweenness values with the values predicted by the baseline model and the proposed model respectively in Figure 4. Note that Figure 4 is a log-log plot due to the skewness of vertex-betweeness distribution. The results show that for the high vertex-betweenness vertices, our model generally provides a better predicted values than the baseline model does.

Next, we compare the goodness-of-fit of the baseline model and the proposed model. In addition, we believe the ranks of the centrality values are important for applications in online game domains. Hence, we also evaluate the persistence of the ranks in the vertex-betweenness values via Kendall's coefficient and Spearman's coefficient.

1) Goodness-of-fit: First, we evaluate the goodness-of-fit of our two models. The *coefficient of determination*, R^2 is the prediction of future outcomes on the basis of given information. R^2 provides a measure of how well the outcomes are likely to be predicted by the model. We assess R^2 shrinkage using 10-fold cross validation and plot the results in Figure 6.

On all eight chat activity networks, the proposed model performs significantly better in terms of the goodness-of-fit than the baseline model. Our model achieves R^2 scores as high as 0.93, and the average is 0.85.

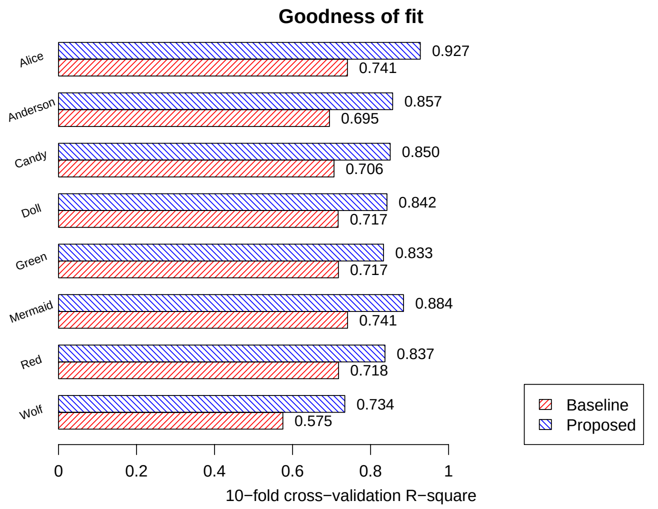


Fig. 6. Comparison of the baseline linear model and the proposed linear model

2) Rank Persistence: To evaluate the number of the highest vertex-betweenness being correctly estimated, we inspect the overlap ratio between the top predicted betweenness vertices and top true betweenness vertices. The estimation exactly captures top $n\%$ high vertex-betweenness vertices if the overlap ratio of top $n\%$ set equals to 1. Table I shows that our proposed model performs better than the baseline model in terms of the capability to capture the highest vertex-betweenness vertices. Normally our approximation reaches the range from 70% to 90% in overlap.

The overlap ratio can expose the number of captured the highest vertex-betweenness vertices, but does not show us how well the rank persist. We therefore use Spearman's coefficient ρ and Kendall's coefficient τ to evaluate the rank persistence. *Spearman's coefficient* is often thought of as being the Pearson correlation coefficient between ranked variables. The n raw scores are first converted to ranks x_i, y_i, and the differences $d_i = x_i - y_i$ between the ranks of each observation on the two variables are calculated. Then, coefficient ρ is given by

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

Kendall's coefficient is a non-parametric statistic used to measure the association or statistical dependence between two measured quantities. More specifically, it is a measure of rank correlation. The coefficient is defined as follows:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)},$$

where n_c is the number of concordant pairs in the data set and n_d is the number of discordant pairs. The denominator can be interpreted as the total number of pairs. A high τ-value means that most pairs are concordant, indicating that the two rankings are consistent.

Generally, users care more about vertices with higher vertex-betweenness values. Therefore, we investigate whether the order of vertices with the top 0.1%, 1%, 5%, 10%, 20%, and 30% vertex-betweenness values are consistent with their estimated order in Figure 5. In our model, the Spearman's coefficients and Kendall's coefficient between the original order and the estimated order are generally acceptable, and much better than the baseline model.

V. Online-Game Social Network Analysis

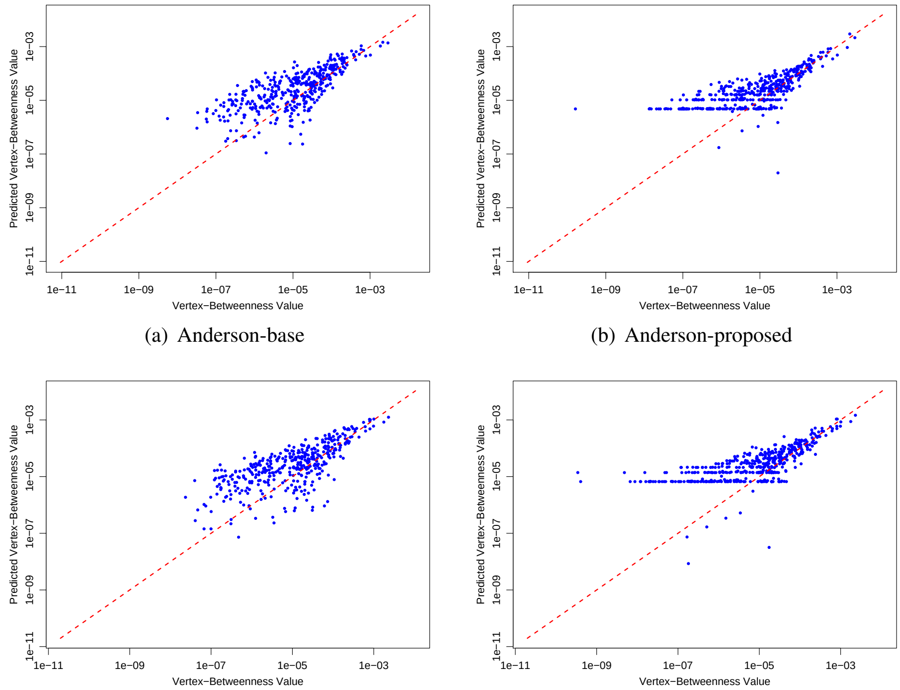
In recent years, researchers have shown increasing interest in online social networks such as MySpace, Flickr, Twitter and other large-scale online communities. In this paper, we focus on networks extracted from massively multiplayer online games. Worldwide, the number of active subscribers to such games increased dramatically from fewer than $10,000$ in 1998 to more than 16 million in 2008.

A. Social Interactions in MMORPGs

Massively multi-player online role-playing games (MMORPGs) are a genre of computer role-playing games in which a large number of players interact with one another in a virtual game world. In this virtual game world, each player assumes the role of a character and controls many of its actions.

According to [8], the social interactions of players are the phenomena that make virtual worlds "sticky" and it is believed that feature enables game worlds to retain long-term players. Most MMORPGs exploit players' social skills and provide various mechanisms to support players' interactions. Social interactions in MMORPGs are typically integrated with the mechanisms of the games. Through the interactions, players may form relationships, varying from simply being a member of a cohesive team, to friendship, or even romance.

Among the numerous interactive mechanisms, *text conversation* is the most widely used. By sending instant text-messages, players can exchange ideas and share information



(a) Anderson-base (b) Anderson-proposed

(c) Green-base (d) Green-proposed

Fig. 4. A sampled view of predicted values vs. true vertex-betweenness values for the baseline model and the proposed model respectively. Note that the x-axis and y-axis are both in log-scale

TABLE I
THE OVERLAP RATIO BETWEEN TOP PREDICTED BETWEENNESS VERTICES AND TOP TRUE BETWEENNESS VERTICES

	Alice		Anderson		Candy		Doll	
	baseline	proposed	baseline	proposed	baseline	proposed	baseline	proposed
Top 0.1%	0.7187500	0.9062500	0.6930693	0.7425743	0.6271186	0.6440678	0.5909091	0.5909091
Top 1%	0.7576687	0.8312883	0.7200791	0.7932740	0.7226891	0.7915966	0.7589286	0.8125000
Top 5%	0.7882497	0.8451652	0.7490607	0.8133281	0.7758737	0.8366935	0.7321747	0.8065954
Top 10%	0.7451820	0.7809728	0.7354424	0.7884330	0.7488661	0.8074920	0.7077300	0.7763422
Top 20%	0.7260630	0.7412053	0.7241720	0.7677706	0.7148497	0.7573492	0.6903542	0.7441524
Top 30%	0.7514021	0.7637402	0.7455264	0.7801285	0.7444569	0.7644457	0.7243633	0.7554021

	Green		Mermaid		Red		Wolf	
	baseline	proposed	baseline	proposed	baseline	proposed	baseline	proposed
Top 0.1%	0.6250000	0.6477273	0.5681818	0.7272727	0.5166667	0.6666667	0.5370370	0.5925926
Top 1%	0.7197740	0.7966102	0.7466368	0.8004484	0.6738411	0.7847682	0.6333333	0.6592593
Top 5%	0.7356062	0.8202755	0.7956989	0.8360215	0.7470199	0.8049669	0.6771566	0.7219548
Top 10%	0.7392482	0.7943334	0.7592385	0.7796193	0.7612978	0.8104618	0.6822136	0.7456968
Top 20%	0.7222191	0.7666347	0.7310492	0.7294816	0.7444343	0.7755524	0.6914963	0.7468308
Top 30%	0.7471312	0.7834381	0.7676172	0.7539564	0.7582345	0.7756690	0.7242613	0.7636173

easily and quickly. Moreover, most MMORPG players make acquaintances and even maintain friendships by exchanging text messages.

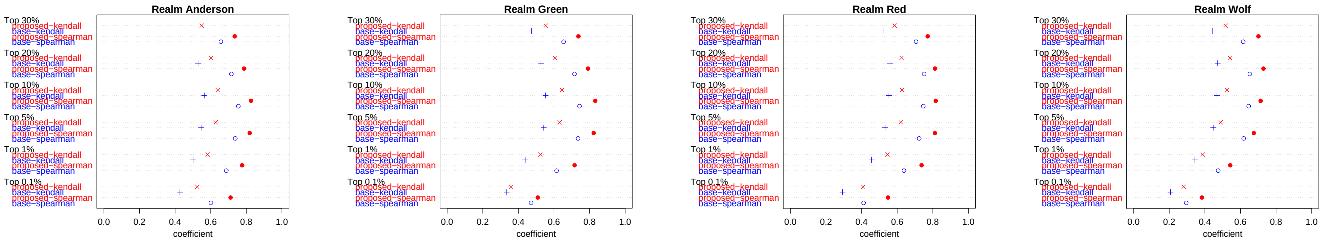
B. In-game demographics

To understand the characteristics of in-game populations, we analyze the demographics of Fairyland Online players. In the game, the gender, race, and appearance of characters are customizable, so players can create their own unique characters. The game offers three playable races: Humans, Elves, and Dwarvens. Figure 7 shows the in-game population by gender and race in the eight realms of Fairyland Online.

First, we observe that there is an imbalance between male and female players. The male to female ratios in the eight realms are: 0.73 in Alice, 0.48 in Anderson, 0.86 in Candy, 0.49 in Doll, 0.48 in Green, 0.82 in Mermaid, 0.70 in Red and 0.44 in Wolf. The male population is almost double the female population in the three realms with largest total population: Anderson, Green, and Wolf.

Second, the Human race is significantly more popular than the other two races. The proportions of the three races (Humans, Elves, and Dwarvens) in the realms are nearly the same: $100:54:32$ in Alice, $100:54:39$ in Anderson, $100:58:28$ in Candy, $100:54:38$ in Doll, $100:50:38$ in Green, $100:58:31$ in Mermaid, $100:59:34$ in Red, and $100:51:33$ in Wolf.

The level of a character is a kind of social capital in MMORPGs. Compared to low-level characters, high-level characters can conquer difficult dungeons more easily, They are also capable of producing more valuable virtual artifacts, and completing quests more quickly. For the most part, low-level characters prefer to seek help from high-level characters.



(a) Anderson (b) Green (c) Red (d) Wolf

Fig. 5. To evaluate the rank persistence of the models, we use $\circ$ and $\bullet$ to identify Spearman's coefficient for the baseline model and the proposed model respectively; and $+$ and $\times$ to identify Kendall's coefficient for the baseline model and the proposed model respectively

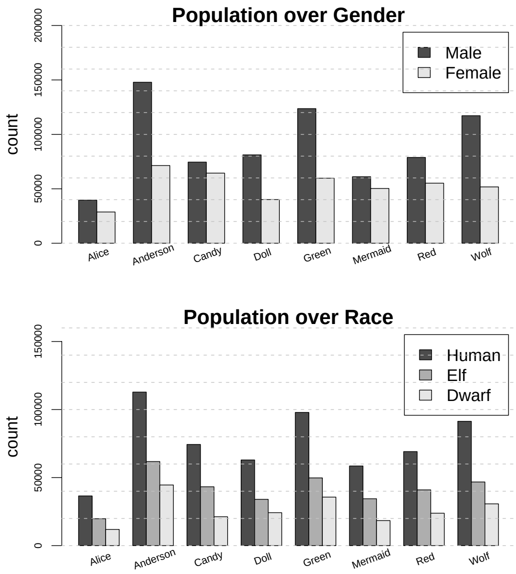


Fig. 7. A snapshot of in-game demographics in terms of gender and race

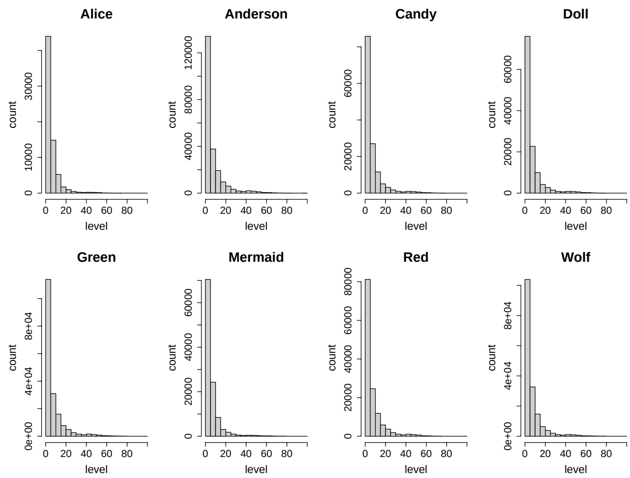


Fig. 8. The distributions of levels in eight realms of Fairyland Online

As a result, high-level characters can build a network of relationships more easily than low level characters.

Figure 8 shows that the level distributions in all 8 realms follow a power-law distribution with an exponential coefficient range of 1.50 to 1.54. This finding corresponds with the Pareto principle (also known as the 80-20 rule), which holds in the virtual game world as well as in the real world.

C. Topology of a Chat Activity Network

Table II summarizes the topological characteristics of the chat activity networks in the eight realms of Fairyland Online. The size of the networks ranges from $32,690$ to $101,150$. The distributions of the in-degrees, out-degrees, and bi-degrees follow a power-law distribution with α equal to 1.37, 1.40, and 1.40 respectively. The mean clustering coefficients of the networks are small. The size of the giant strongly connected components (gscc) ranges from 66.41% to 79.24%, and the diameter ranges from 11 to 14.

1) Degree distribution: We begin with the degree distribution in MMORPG social networks. Because chatting is a directional edge from one entity to another, the degree of vertices can be classified into two categories: *in-degree* and *out-degree*. For a vertex v, we denote the vertices linked to v as *in-degree neighbors* and the vertices from v as *out-degree neighbors*. These two kinds of degrees represent different types of activity; the out-degree of a vertex represent how often an individual's contacts other players, while the in-degree represents how frequently the individual is contacted by other players.

Figure 9(a) shows that the out-degree distributions of the four largest chat activity networks follow a power law distribution; and Figure 9(b) shows that the in-degree distributions are also power law distributions. The in-degree and out-degree distributions look similar, which implies that active individuals coincide with popularity distribution in such instant text-conversation.

2) Clustering coefficient: A remarkable property of most real-world scale-free networks is the high clustering coefficient of their vertices. The clustering coefficient of a vertex in a graph quantifies how close its neighbors are to being a clique (a complete graph). The measurement was introduced by Watts and Strogatz [6]. The clustering coefficient for undirected graphs is defined as follows:

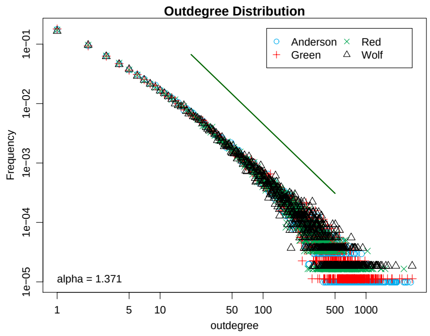
$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i-1)}, \text{where } e_{ij}, e_{ik}, e_{jk} \in E(G).$$

In this work, we convert chat activity networks into undirected graphs to calculate the clustering coefficients. For the four largest realms: Anderson, Green, Red, and Wolf, the mean clustering coefficients are 0.10567919, 0.10201802,

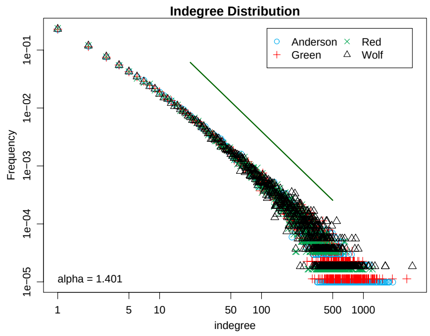
TABLE II
SUMMARY OF CHAT ACTIVITY NETWORKS IN FAIRYLAND ONLINE

Realm	no. of vertices	no. of edges	avg. degree	avg. cc	ratio of gscc	avg. degree in gscc	diameter in gscc	avg. path len. in gscc
Alice	32,690	445,528	27.25	0.0847	66.41%	39.05	12	3.70
Anderson	101,150	2,363,864	46.74	0.1057	79.24%	58.06	11	3.74
Candy	59,534	1,209,526	40.63	0.0938	74.01%	53.65	12	3.70
Doll	44,891	876,911	39.07	0.1069	73.83%	51.69	12	3.73
Green	88,599	2,073,034	46.80	0.1020	78.70%	58.50	13	3.73
Mermaid	44,656	684,145	30.64	0.0805	67.96%	43.34	14	3.69
Red	60,418	1,239,473	41.03	0.0976	76.88%	52.32	13	3.74
Wolf	54,039	1,233,609	45.66	0.1039	75.86%	59.04	12	3.72

cc: clustering coefficient, gscc: giant strongly connected component



(a) Outdegree



(b) Indegree

Fig. 9. Degree distribution of chat-activity network in four realms: Anderson, Green, Red, and Wolf

0.09769602, and 0.10387252. In Figure 10, we plot the clustering coefficients distributions of the four realms. More than 80% of the vertices have clustering coefficients less than 0.2. The distributions are skewed. To provide some extra information, we also plot the average clustering coefficients of vertices grouped based on their out-degrees in Figure 11. There is an obvious correlation between the degree and the clustering coefficient. The clustering coefficients of vertices with out-degrees less than 50 in a chat activity network decline exponentially as their degree increases. For degree 50 and above, the relation between the degree and the clustering coefficients becomes weaker. However, we observe that there are some high degree vertices still remain relative high clustering coefficients.

VI. RELATED WORK

Social network analysis dates back to the late 1970s. Freeman [1] proposed the concept of "betweenness centrality" as a structural property of social networks. The definition of betweenness centrality was relaxed by Newman [9] in 2005 to count how often a node is traversed during a random walk between two other nodes. However, the measurement proposed by Freeman is widely used in the analysis of social networks [10]–[12].

Determining exact vertex-betweenness centrality is computationally-expensive. The currently fastest-known algorithm proposed by Brande [2] requires $\mathcal{O}(|V||E|)$ time. Based on this algorithm, Brande [13] provides several efficient algorithms for computing the shortest-path betweenness variants including edge-betweenness centrality, stress centrality, load centrality, and etc..

Trying to reduce the cost of vertex-betweenness computation, Bader et al. [14] present an approximating algorithm based on an adaptive sampling technique, which reduces the number of single-source shortest path computations for vertices with high centrality values. Nevertheless, the time complexity of the approximation is still bounded by $\mathcal{O}(|V||E|)$.

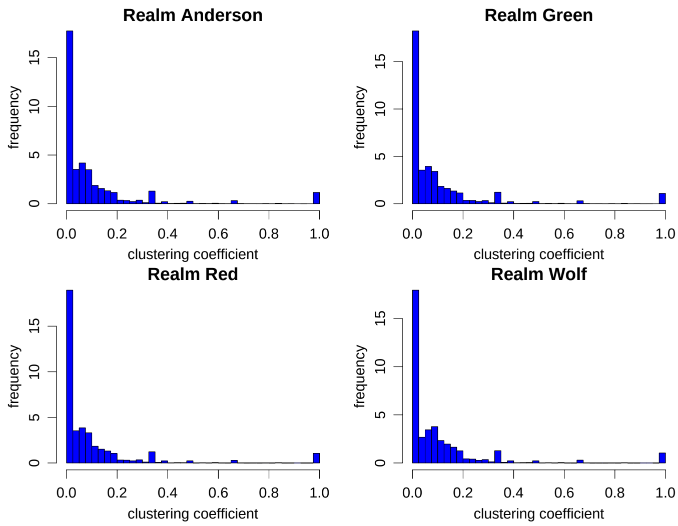


Fig. 10. Clustering coefficient distribution in the four largest realms: Anderson, Green, Red, and Wolf

The social networks in online games have attracted a great deal of attention in recent years. Jakobsson [15] discussed the formation of social communities in a classic online game called Ever Quest. For the numerous kinds of interactions in an online game, such as healing and purchasing, Ducheneaut et al. [16] outlined different patterns of interactivity and discussed how they are affected by the structure of the game. They

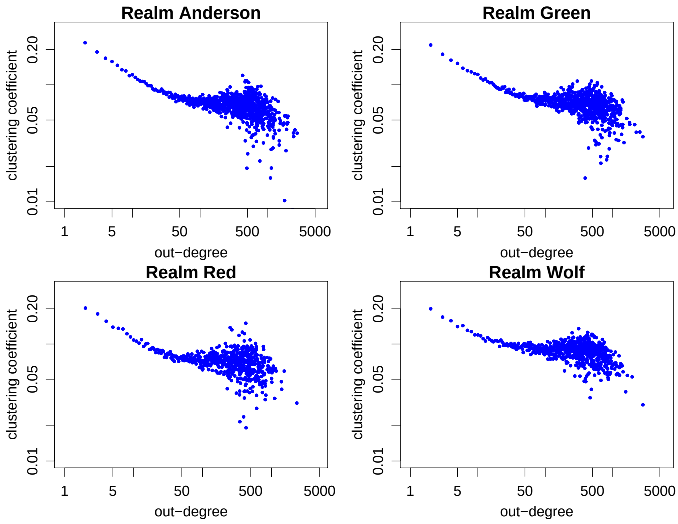


Fig. 11. Clustering coefficient dependence of degree in the four largest realms: Anderson, Green, Red, and Wolf

also made a series of recommendations for the design of social activities in MMORPGs. Chen et al. [17] found that participants who have a higher degree of social interaction tend to play much longer, and players who are closer in network topology tend to team up for longer periods. Moreover, to investigate the evolution of social communities in online games, Ducheneaut et al. [18] used longitudinal data collected directly from the popular online game World of Warcraft to identify the grouping patterns of players.

To identify influential players, Kirman [19] analyzed the online game called "Familiar" to identify Hardcore players. Among the various classifications in [20], the players can be classified into two classes: *Hardcore* players and *Casual* players. Kirman classified the players according to the number of in-game activities. However, the data set only involved 157 active users and recorded 1546 distinct interactions between players. To our knowledge, this paper is the first paper that reports studies on the influential player analysis for large online game social networks. We were able to identify high betweenness nodes in networks of this large scale thanks to the estimation method we have designed.

VII. Conclusion

In this paper, we discuss how that dynamic information about degree and transitivity can help us understand the evolution of social networks. Moreover, based on the local dynamic information, we have proposed a fast and accurate method for estimating vertex-betweenness values in a real-world application; the method identifies influential individuals in MMORPGs. To shed more light on the social interactions between MMORPG players, we examined the demographics of Fairyland Online players and identified the structural characteristics of chat activity networks based on private text-conversations.

Compared to the the Brandes algorithm, our estimation requires $\mathcal{O}(b^2|V|)$ time, where b is the average degree in the network. In Fairyland Online, the average value of b is approximately 50 while the average value of $|E|$ is $1,265,761$. In other words, the proposed model is 506 times faster in terms of time complexity. We evaluate the goodness-of-fit by shrinkage R^2 using 10-fold cross validation, and the rank persistence by Spearman's coefficient and Kendall's coefficient with promising results.

VIII. Acknowledgement

This work was supported in part by Taiwan E-learning and Digital Archives Programs (TELDAP) sponsored by the National Science Council of Taiwan under the grants NSC99-2631-H-001-018 and NSC99-2631-H-001-020. It was also supported in part by the National Science Council of Taiwan under the grants NSC98-2221-E-001-017.

References

[1] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, March 1977. [Online]. Available: http://links.jstor.org/sici?sici=0038-0431%28197703%2940%3A1%3C35%3AASOMOC%3E2.0.CO%3B2-H

[2] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of Mathematical Sociology*, vol. 25, pp. 163–177, 2001.

[3] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," in *Proceedings of the twelfth international conference on Information and knowledge management*. ACM New York, NY, USA, 2003, pp. 556–559.

[4] M. Newman, "The structure and function of complex networks," *Arxiv preprint cond-mat/0303516*, 2003.

[5] B. Woodcock, "MMOGCHART.COM," http://www.mmogchart.com/.

[6] D. Watts and S. Strogatz, "Collective dynamics of small-worldnetworks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[7] T. Schank and D. Wagner, "Approximating clustering coefficient and transitivity," *Journal of Graph Algorithms and Applications*, vol. 9, no. 2, pp. 265–275, 2005.

[8] R. Bartle, *Designing virtual worlds*. New Riders Games, 2003.

[9] M. Newman, "A measure of betweenness centrality based on random walks," *Social networks*, vol. 27, no. 1, pp. 39–54, 2005.

[10] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the National Academy of Sciences*, vol. 101, no. 11, p. 3747, 2004.

[11] D. Lusseau and M. Newman, "Identifying the role that animals play in their social networks," *Proceedings: Biological Sciences*, vol. 271, pp. 477–481, 2004.

[12] P. Holme, B. Kim, C. Yoon, and S. Han, "Attack vulnerability of complex networks," *Physical Review E*, vol. 65, no. 5, p. 56109, 2002.

[13] U. Brandes, "On variants of shortest-path betweenness centrality and their generic computation," *Social Networks*, vol. 30, no. 2, pp. 136–145, 2008.

[14] D. Bader, S. Kintali, K. Madduri, and M. Mihail, "Approximating betweenness centrality," *Algorithms and Models for the Web-Graph*, pp. 124–137, 2007.

[15] M. Jakobsson and T. Taylor, "The Sopranos meets EverQuest: social networking in massively multiplayer online games," in *Proceedings of the 2003 Digital Arts and Culture (DAC) conference, Melbourne, Australia*, 2003, pp. 81–90.

[16] N. Ducheneaut and R. Moore, "The social side of gaming: a study of interaction patterns in a massively multiplayer online game," in *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. ACM New York, NY, USA, 2004, pp. 360–369.

[17] K.-T. Chen and C.-L. Lei, "Network game design: Hints and implications of player interaction," in *Proceedings of ACM NetGames 2006*, Singapore, Oct 2006.

[18] N. Ducheneaut, N. Yee, E. Nickell, and R. Moore, "Alone together?: exploring the social dynamics of massively multiplayer online games," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, p. 416.

[19] B. Kirman and S. Lawson, "Hardcore Classification: Identifying Play Styles in Social Games Using Network Analysis," in *Proceedings of the 8th International Conference on Entertainment Computing*. Springer, 2009, p. 251.

[20] R. Bartle, "Hearts, clubs, diamonds, spades: Players who suit MUDs," *Journal of MUD research*, vol. 1, no. 1, p. 19, 1996.