Emergence of Fractals in Social Networks: Analysts of Community Structure and Interaction Locality

<table>
<thead>
<tr>
<th>著者別名</th>
<th>津川 翔 □ 大崎 博之</th>
</tr>
</thead>
</table>

doi: 10.1109/COMPSAC.2014.80
Emergence of Fractals in Social Networks: Analysis of Community Structure and Interaction Locality

Sho Tsugawa  
Faculty of Engineering, Information and Systems  
University of Tsukuba  
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan  
Email: s-tugawa@cs.tsukuba.ac.jp

Hiroyuki Ohsaki  
School of Science and Technology  
Kwansei Gakuin University  
2-1 Gakuen, Sanda, Hyogo 669-1337, Japan  
Email: ohsaki@kwansei.ac.jp

Abstract—Research on social network analysis (SNA) has been actively pursued. Most SNAs focus on either social relationship networks (e.g., friendship and trust networks) or social interaction networks (e.g., email and phone call networks). It is expected that the social relationship network and social interaction network of a group would be closely related to each other. For instance, people in the same community in a social relationship network are expected to communicate with each other more frequently than with people in different communities. To the best of our knowledge, however, there is not yet any empirical evidence to support the existence of such interaction locality in large-scale online social networks. This paper aims to bridge the evidence gap between intuition about interaction locality and confirmation that it occurs. We investigate the strength of interaction locality in large-scale social networks by analyzing several types of data: logs of mobile phone calls, email messages, and message exchanges in a social networking service. Our results show that strong interaction locality is observed equally in the three datasets and suggest that the strength of the interaction locality is fractal, by which we mean that the strength is invariant with regard to the scale of the community.

Keywords—social network; interaction locality; community; fractal

I. INTRODUCTION

Research on social network analysis (SNA) has been actively pursued [1-3]. In SNA, individuals are represented by nodes in a graph, and social ties among them are represented by links [1-3]. The resulting graph is then analyzed to understand complex social phenomena, which involve interactions among a large number of people.

Most SNAs focus on either social relationship networks [4-8] or social interaction networks [9-14]. Links in social relationship networks represent relationships among individuals, such as friendship or trust. In contrast, links in social interaction networks represent actual interactions between individuals, such as email communication, phone calls, or face-to-face conversation.

Although it is intuitively expected that the social relationship network and social interaction network of a group would be closely related to each other, there is not much evidence on the relationship between these two types of networks. For instance, it is expected that the community structure of a social relationship network will be related to interactions among the people in the network. A community in a social relationship network is a densely connected subgraph, and in many cases, it represents a group in the real world [15]. Therefore, people in the same community in a social relationship network are expected to communicate with each other more frequently than with people in different communities. We call this characteristic interaction locality (Fig. 1). The concept of the interaction locality is not new, since it has been studied in the area of social sciences [16-18]. However, empirical studies are limited to analysis on small-scale social networks in offline environments. In contrast, recent trends in SNAs are shifting from small-scale analysis in offline environments to large-scale analysis in online environments [1, 19]. Better understanding of interaction locality in large-scale social networks in offline and online environments should be useful both for understanding social phenomena and for developing novel services, such as inferring potential communication demands and detecting anomalous interactions within a social relationship network by SNA. To the best of our knowledge, however, there is not yet any empirical evidence to support the existence of interaction locality in large-scale online social networks because traditional SNAs focus on small-scale social networks in offline environments, and in recent SNAs focusing on online environments, social relationship networks and interaction networks have, in most case, been analyzed independently of each other.

This paper aims to bridge the evidence gap between intuition about interaction locality and confirmation that it occurs. We therefore investigate the strength of interaction locality in large-scale social networks by analyzing data on mobile phone calls [20], email messages [21], and message exchanges in a social networking service (SNS) [22]. We obtain communities of users of mobile phones, email, and an SNS by analyzing actual or inferred friendship networks, which are one type of social relationship network. As a measure of the strength of interaction locality, we choose the ratio of the number of interactions within the community.
to the total number of all interactions.

Our main contributions are summarized as follows.

- We empirically show that 80–90% of interactions are communications among individuals in the same community, and strong interaction locality is observed equally in data on mobile phone calls, email, and SNS messaging.
- We analyze the dynamics of the interaction locality, and show that the strength of interaction locality does not change frequently, over time.
- We find that strength of interaction locality is fractal, by which we mean that the strength is invariant with regard to the scale of the community.
- We show that the community structure itself is also fractal, and this should be one of major reasons that a fractal relation emerges in interaction locality.

The remainder of this paper is organized as follows. In Section II, we introduce works related to social relationship and interaction network analyses. In Section III, we extensively investigate the strength of interaction locality using data on mobile phone calls, email messages, and message exchanges in an SNS. Finally, Section IV concludes this paper and discusses future work.

II. RELATED WORK

In the area of social sciences, the interaction locality has been studied [16-18]. For instance, Lomi et al. [16] propose a model of organizational evolution, in which global dynamics emerge from local interaction among individual organizations. This model succeeded to explain the organizational evolution, which implies the existence of the interaction locality. Leifer [17] discusses the relation between local actions and the roles of actors. Mandel [18] analyzes the local interaction of individuals and their role in a community. These traditional SNAs focus on small-scale social networks in offline environment. In contrast, we investigate the interaction locality in large-scale social networks including online social networks using several types of communication logs.

Analysis of the relationship between social interaction and relationship networks is a new research topic. Eagle et al. [20] analyze the relation between a phone call network of students and faculty at a university and their friendship network. The research reveals that a friendship network structure can be inferred from mobile phone log data. Golder et al. [23] show that in a popular SNS, Facebook, approximately 90% of messages are exchanges between friends. Weng et al. [24] analyze the information diffusion in social networks. They analyze information diffusion network and follow network evolution in Twitter, a popular microblogging service. Consequently, it is shown that information diffusion affects the follow network evolution. The results of these works suggest that social relationship and interaction networks are closely related to each other. However, the interaction locality, which we focus on in this paper, is not yet well understood.

One notable exception to the dearth of research on interaction locality in large-scale social networks is the analysis of a society-wide phone call network performed by Palla et al. [25]. They showed that phone call durations between individuals in the same community are 5.9-fold the durations of calls to community nonmembers. This result strongly suggests the existence of interaction locality. However, the communities in that work were obtained from the phone call network, which is a social interaction network. In contrast, we focus on the communities obtained from social relationship networks.

III. EXPERIMENT

A. Datasets and Methodology

We investigate the strength of interaction locality by using three datasets, which contain a history of interactions and are grouped by medium: mobile phone calls (MIT dataset) [20], email messages (Enron email dataset) [21], and message exchanges on an SNS (Facebook dataset) [22]. The MIT dataset contains mobile phone call logs of students and faculty at MIT (Massachusetts Institute of Technology) from September 2004 to June 2006. The Enron email dataset contains logs on messaging among users of Facebook, which is a popular SNS, during January 2008 to December 2008. The MIT and Facebook datasets also contain “friendship” information among mobile phone and Facebook users, respectively.

We obtained communities (i.e., densely connected clusters) of mobile phone, email, and SNS users from these datasets. Since the MIT dataset and the Facebook dataset
contain friendship information among the users, we obtained communities by applying a popular community detection algorithm called the fast Newman algorithm [26] to the friendship networks constructed from the friendship information. The friendship network is an unweighted undirected network where nodes correspond to users. A link between nodes $i$ and $j$ (corresponding to users $i$ and $j$) is generated when users $i$ and $j$ are friends. Although the Enron email dataset contains neither friendship information nor explicit community information, we inferred communities in the Enron Corporation by using an inferred friendship network, constructed from the history of email communications. The inferred friendship network is an unweighted undirected network where nodes correspond to Enron employees. A link between nodes $i$ and $j$ is generated when the number of email exchanges between employees $i$ and $j$ exceeds 5. We then obtained the communities of the Enron employees from this inferred friendship network by the fast Newman algorithm, and the observation duration of communications for each dataset.

We investigated the strength of interaction locality at two levels: the network level (macroscopic locality) and the community level (microscopic locality). We calculated the strength of macroscopic locality as the ratio of the number of interactions between users in the same community to the number of all interactions. We calculated the strength of the microscopic locality of community $k$ as the ratio of the number of interactions between users in community $k$ to the number of interactions involving users in community $k$. Using these measures, we investigated the strength of interaction locality in each dataset. Note that we used only those interactions for which both ends of the interaction are in the friendship network.

### B. Macroscopic Analysis

We first perform macroscopic analysis, in which we investigate the strength of macroscopic locality by using all communications data in the three datasets. Figure 3 shows the ratio of the number of intracommunity interactions to the number of all interactions (i.e., the strength of macroscopic locality) in each dataset.

---

**Table I**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of users</th>
<th>Number of communities</th>
<th>Observation duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT</td>
<td>60</td>
<td>10</td>
<td>9 months</td>
</tr>
<tr>
<td>Enron</td>
<td>149</td>
<td>9</td>
<td>24 months</td>
</tr>
<tr>
<td>Facebook</td>
<td>63,731</td>
<td>771</td>
<td>12 months</td>
</tr>
</tbody>
</table>
Figure 3 shows that 80–90% of interactions are intracommunity in all datasets. As we intuitively expected, strong interaction locality is observed in mobile phone call, email, and SNS messaging logs. It should be noted that strong locality is commonly observed in communication by various media. This result offers strong evidence to support our intuition.

C. Stationarity Analysis

Communication patterns among users may change over time, which raises a question: does the strength of interaction locality change over time?

To answer the above question, we next investigate the stationarity of the strength of macroscopic locality. The time evolutions of the strength of macroscopic locality and number of interactions by month are shown in Figs. 4 and 5, respectively.

Although Fig. 5 shows that the frequency of interactions varies from month to month, Fig. 4 shows that the strength of macroscopic locality is quite stable in all datasets. Mean strengths of all monthly macroscopic localities in MIT, Enron, and Facebook are 0.94, 0.85, and 0.90, respectively. Standard errors of them are 0.008, 0.02, and 0.005, respectively. This suggests that the strength of interaction locality does not change frequently, even over time. An explanation for this is that neither the community structure in social relationship networks nor the communication patterns of the users change frequently. Longer periods of observation may lead to a different conclusion, but changes in the interaction locality are small for the three datasets here.

D. Microscopic Analysis

We next investigate interaction locality from the microscopic viewpoint rather than from the macroscopic viewpoint. We investigate the strength of microscopic locality for each obtained community. Because the number of communities in the Facebook dataset is large (see Tab. I), we focus on the MIT dataset and the Enron dataset in this subsection.

Table II shows the community size, the number of intracommunity interactions, the number of intercommunity interactions, and the strength of microscopic locality for each community in the MIT dataset. Table III shows the analogous values from the Enron dataset.

From Table II, strong interaction locality is observed in most of the communities in the MIT dataset. In some communities (in particular communities 7, 9, and 10), interaction locality is very weak. This is likely because the sizes of the communities are too small. Table III shows that strong interaction locality is observed in the Enron dataset also.

We find from these results that the strength of microscopic locality is similar between differently sized communities. Moreover, we note that when we consider the Enron Corporation as a community, the strength of macroscopic locality is approximately 0.8. Table III shows that when we partition the set of Enron employees into several communities, the strength of interaction locality in those communities is also approximately 0.8.

This reminds us of a fractal structure. In particular, this result suggests that the strength of interaction locality is independent from the scale of the community. Many social networks have hierarchical community structures [27].
Similar interactional locality strengths may be observed at different levels of the hierarchy. We further investigate this phenomenon in the next section by using a large-scale dataset with hierarchical communities, the Facebook dataset.

E. Effect of the Scale of Community

To investigate the effects of the scale of community, we obtained communities of different scales and examined the strength of microscopic localities in these communities. An illustrative example of obtaining communities of different scales is shown in Fig. 6. To obtain communities of different scales, we simply used the community detection algorithm [26] recursively. We recursively partition a community in the Facebook dataset into smaller communities. To do so, we first obtain communities from the Facebook dataset. For each community, we apply the community detection algorithm to the subgraph corresponding to that community and calculate the strength of interaction locality in the community. More specifically, for subgraph $G_i = (V_i, E_i)$, where $V_i$ is the set of nodes belonging to community $i$ and $E_i$ is the set of links connecting nodes in $V_i$, we divide $G_i$ into $n$ communities. We then calculate the strength of microscopic locality in community $k$ ($1 \leq k \leq n$). Note that we used only those interactions for which both ends of the interaction are in $V_i$ when calculating the strength of microscopic locality in community $i$. We recursively repeat this procedure until the size of the community falls below 100.

Figure 7 shows the relation between the size of a community and the strength of microscopic locality in that community.
These results show that although the sizes of the communities differ, most of the communities have similar strengths of interaction locality (around 0.8). The mean of the strength of microscopic locality is 0.80. Standard error of it is 0.006. This result supports our hypothesis: interaction locality is fractal. We can observe similar strengths in interaction locality at different scales of community. We note that the strength of interaction locality in smaller communities varies more widely than in larger communities. The cause of this may be that community detection is too aggressive and thus artificially splits communities.

F. Cause of the Fractal Pattern

In this section, we discuss why interaction locality is fractal. In the Facebook dataset, most of the messages (97% of messages) are exchanged between friends, and so the strength of interaction locality is almost completely determined by the ratio of intracommunity links to all links. Therefore, a simple and intuitive explanation for the emergence of a fractal pattern in interaction locality is that the community structure itself is fractal (i.e., the ratio of intracommunity links to all links is insensitive to the scale of the communities). Ferrara [28] shows that communities in the Facebook are similar to each other. We therefore expect that community structure in Facebook is fractal.

We investigate the ratio of intracommunity links to all links in each community. Figure 8 shows the relation between the size of a community and the ratio of intracommunity links to all links in the community.

This result shows that although the sizes of the communities are different, the average of the ratio of intracommunity links to all links in the community is around 0.8. The mean of the ratio of intracommunity links to all links in the community is 0.79. Standard error of it is 0.005.

This result supports our hypothesis that community structure is fractal (i.e., the ratio of intracommunity links to all links is invariant with respect to the scale of the community). Therefore, we suggest that this is the main cause of the emergence of a fractal pattern in interaction locality.

G. Practical Implications

The fact that strong interaction locality exists may be useful in designing new services, such as anomalous communication detection [29], traffic engineering [30], and virtual network embedding [31]. Our results show the existence of strong interaction locality, which implies that a large amount of communication between members of different communities is anomalous. Our results also show that the strength of interaction locality does not change frequently, which implies that rapid changes in the strength of interaction locality can identify anomalous events. Using these characteristics, a novel scheme for anomaly detection can be designed. Moreover, our results imply that coarse-grained traffic demands of individuals can be inferred from their communities. Such traffic pattern is expected to be useful for traffic engineering and virtual network embedding.

The fractal pattern in interaction locality is also useful: if we can know the strength of interaction locality in some small communities, then we can also know the strength of interaction locality in larger communities. It is not feasible to observe a society-wide social relationship network and investigate the strength of interaction locality in each community, but our results suggest that observing a fraction of a large-scale social network is enough to know the strength of interaction locality in the network.
IV. CONCLUSION AND FUTURE WORK

This paper has empirically validated the hypothesis that people in the same community in a social relationship network communicate with each other more frequently than with people in different communities. This provides empirical evidence to support the intuition that interaction locality exists in large-scale social networks. Our results support our hypothesis, and we have shown that 80–90% of interactions occur within communities in mobile phone, email, and SNS messaging communications. Moreover, our results suggest that the strength of interaction locality is fractal (i.e., the strength is invariant with regard to the scale of the community). We also showed that the community structure itself is fractal and that this should be the main cause of the emergence of a fractal pattern in interaction locality.

In the future, we plan to design novel services on the basis of the fact that strong interaction locality exists. For instance, anomalous interaction detection, traffic engineering, and virtual network embedding should be promising applications.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 25280030.

REFERENCES


