Electronic Preprint for Journal of Information Processing Vol.23 No.4

Invited Paper

Community Structure and Interaction Locality in Social Networks

Sho Tsugawa[1,a)] Hiroyuki Ohsaki[2,b)]

Received: December 3, 2014, Accepted: March 20, 2015

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 should 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 much understanding on such interaction locality in large-scale online social networks. This paper aims to bridge the gap between intuition about interaction locality and empirical evidences observed in large-scale social networks. We investigate the strength of interaction locality in large-scale social networks by analyzing different 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 strength of the interaction locality is invariant with regard to the scale of the community. Moreover, we discuss practical implications as well as possible applications.

Keywords: social network, interaction locality, community, anomaly detection

1. Introduction

Research on social network analysis (SNA) has been actively pursued [1], [2], [3]. In SNA, individuals are represented by nodes in a graph, and social ties among them are represented by links [1], [2], [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], [5], [6], [7], [8] or *social interaction networks* [9], [10], [11], [12], [13], [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 should be closely related to each other, there is not much understanding on the relationship between these two types of networks. For instance, it is expected that the community structure of a social relationship network should affect the 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 a 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], [17], [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 both offline and online environments should be useful for understanding social phenomena and also 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 that proves 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 cases, been analyzed independently of each other.

This paper aims to bridge the gap between intuition about interaction locality and empirical evidences in large-scale social networks. 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 social relationship

[1] Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Ibaraki 305–8573, Japan
[2] School of Science and Technology, Kwansei Gakuin University, Sanda, Hyogo 669–1337, Japan
[a)] s-tugawa@cs.tsukuba.ac.jp
[b)] ohsaki@kwansei.ac.jp

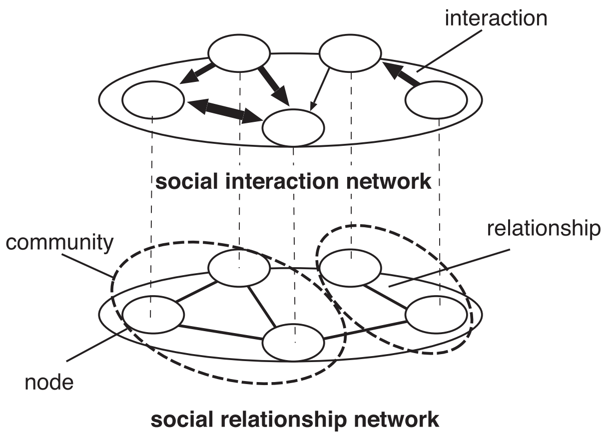


Fig. 1 Intuition about the interaction locality. People in the same community in a social relationship network are expected to communicate with each other more frequently than with people in different communities.

networks. 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 find that the strength of interaction locality is invariant with regard to the scale of the community.
- We analyze the dynamics of the interaction locality, and show that the strength of interaction locality does not change frequently over time.
- We introduce possible applications of our findings and also show the feasibility of one of those applications through preliminary experiments.

The remainder of this paper is organized as follows. In Section 2, we introduce past studies related to social relationship and interaction network analyses. In Section 3, we extensively investigate the strength of interaction locality using data on mobile phone calls, email messages, and message exchanges in an SNS. Section 4 discusses practical implications of our results. Section 5 introduces a possible application of our findings, and examines its feasibility through case studies. Finally, Section 6 concludes this paper and discusses future works.

2. Related Work

In the area of social sciences, the interaction locality has been studied [16], [17], [18]. For instance, Lomi et al. [16] propose a model of organizational evolution, in which global dynamics emerges 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 environments. 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 rather new research topic. Eagle et al. [20] analyze the relation between a phone call network of students and faculty at a university and the friendship network among them. The research reveals that the structure of a friendship network can be inferred from mobile phone log data. Golder et al. [23] show that in a popular SNS, Facebook, approximately 90% of messages are exchanged among friends. Weng et al. [24] analyze the information diffusion in social networks. They analyze information diffusion network and evolution of follow network in a micro-blogging service. Consequently, it is shown that information diffusion affects the evolution of follow network. 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.

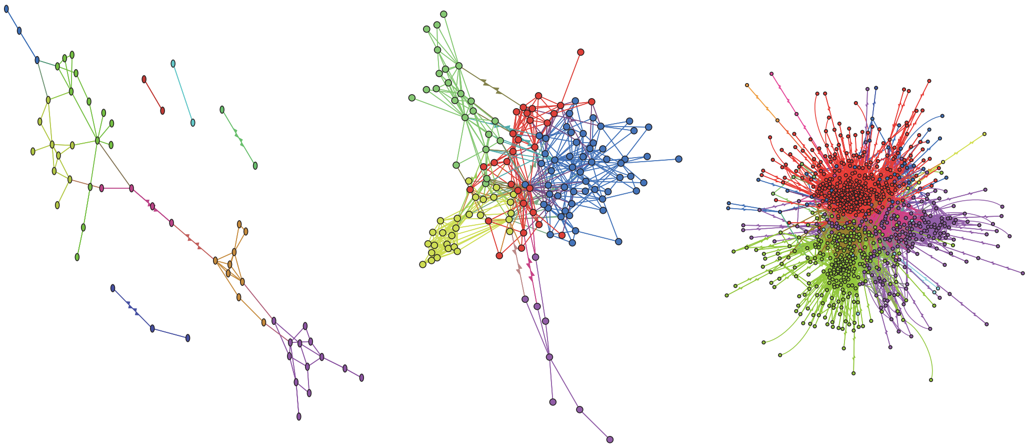
One notable exception to the dearth of research on interaction locality in large-scale social networks is the analysis of a society-wide phone call network performed by Palla et al. [25]. They show that durations of phone calls between individuals in the same community are 5.9-fold the durations of calls between individuals in a different community. This result strongly suggests the existence of interaction locality. However, the communities in Ref. [25] 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.

3. Experiment

3.1 Datasets and Methodology

We investigate the strength of interaction locality using three datasets, which contain records of interactions on different communication media: mobile phone calls (MIT dataset) [20], email messages (Enron dataset) [21], and message exchanges on an SNS (Facebook dataset) [22]. The MIT dataset contains mobile phone call logs among students and faculty at MIT (Massachusetts Institute of Technology) from September 2004 to June 2006. The Enron dataset contains logs of emails sent to and from employees of the Enron Corporation during April 2000 to March 2002. The Facebook dataset contains logs on messaging among users in 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 directly applying a popular community detection algorithm called the fast Newman algorithm [26] to the friendship networks constructed from the friendship information. A 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. Since the Enron dataset contains neither friendship information nor explicit community information, we inferred communities in the Enron dataset based

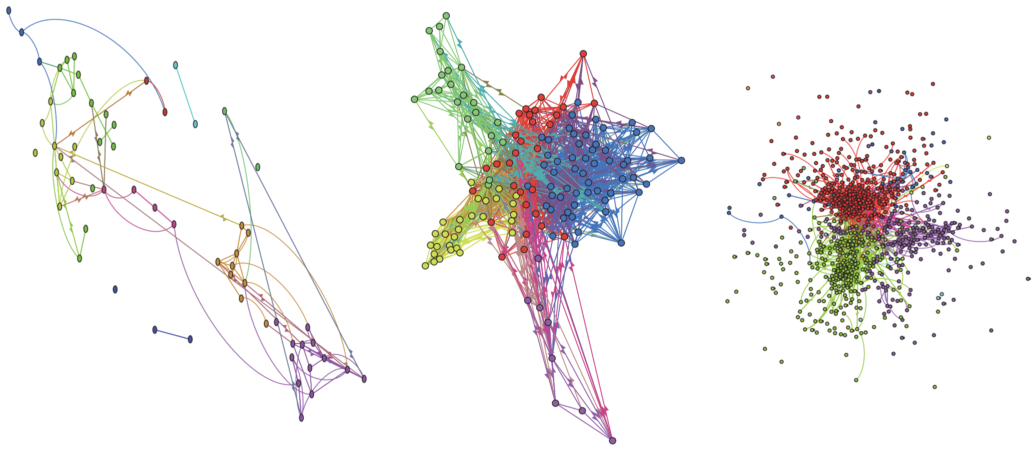


(a) MIT dataset

(b) Enron dataset

(c) Facebook dataset (A subnetwork, where nodes belong to the largest community, is shown. Communities obtained from the subnetwork are depicted in different colors.)

Fig. 2 Visualization of friendship networks in three datasets (communities detected with the Fast Newman algorithm are depicted in different colors).



(a) MIT dataset

(b) Enron dataset

(c) Facebook dataset

Fig. 3 Visualization of interaction networks in three datasets (communities detected with the Fast Newman algorithm are depicted in different colors). A link between two nodes represents the existence of interaction between them. These figures visually show that most interactions occur within communities.

Table 1 Overview of datasets.

dataset	number of users	number of communities	observation duration
MIT	60	10	9 months
Enron	149	9	24 months
Facebook	63,731	771	12 months

on an inferred friendship network, which was constructed from the records 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 a predefined threshold τ. We used $\tau = 5$ in our experiments. We then obtained the communities from this inferred friendship network using the fast Newman algorithm [26]. **Figure 2** is visualization of the friendship networks and communities observed in the MIT dataset, the Enron dataset, and the Facebook dataset. **Figure 3** is visualization of the interaction networks in three datasets. **Table 1** summarizes several properties of three datasets: the number of users, which correspond to nodes in friendship networks, the number of communities detected by the fast Newman algorithm, and the observation duration of communications.

We investigate the strength of interaction locality at three levels: the network level (macroscopic locality), the community level (mesoscopic locality), and the node level (microscopic locality). Let i be a user, c_i be a community to which user i belongs, V_c be a set of users in community c, $n_{i,j}$ be a number of interactions from user i to user j. An interaction from user i to user j corresponds to a phone call originated from user i to user j, an email message from user i to user j, and a message from user i to user j in the MIT dataset, the Enron dataset, and the Facebook dataset, respectively. We define the number of intra-community

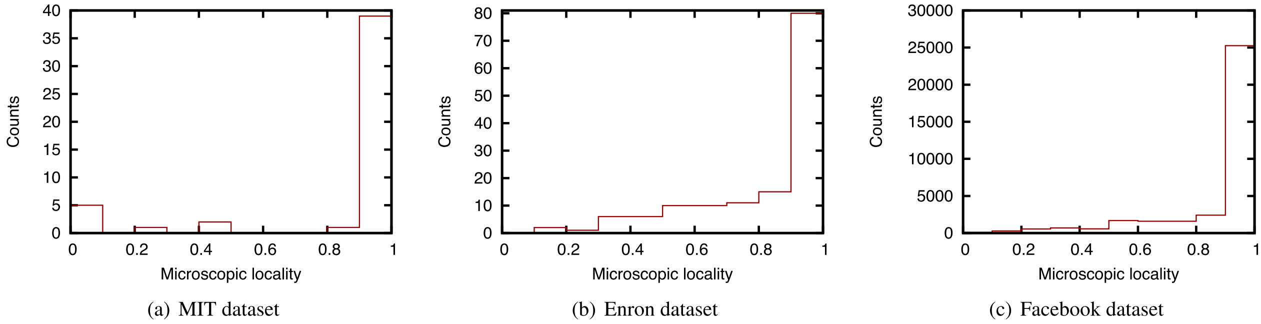


Fig. 4 Distributions of the strength of microscopic locality (Counts shown in the vertical axis is defined as the number of users whose strengths of microscopic locality are in the range of each bin).

Table 2 The strength of the macroscopic locality: Ratio of the number of intra-community interactions to the number of all interactions.

MIT	Enron	Facebook
0.93	0.82	0.85

interaction of user i, N_i^{intra}, and the number of inter-community interaction of user i, N_i^{inter} as

$$N_i^{\text{intra}} = \sum_{j \in V_{c_i}} n_{i,j}, \quad (1)$$

$$N_i^{\text{inter}} = \sum_{j \notin V_{c_i}} n_{i,j}. \quad (2)$$

Then, the strength of macroscopic locality of friendship network G, the strength of microscopic locality of user i, and the strength of mesoscopic locality of community c are defined as

$$\rho^{\text{macro}}(G) = \frac{\sum_{i \in V} N_i^{\text{intra}}}{\sum_{i \in V}(N_i^{\text{intra}} + N_i^{\text{inter}})}, \quad (3)$$

$$\rho^{\text{micro}}(i) = \frac{N_i^{\text{intra}}}{N_i^{\text{intra}} + N_i^{\text{inter}}}, \quad (4)$$

$$\rho^{\text{meso}}(c) = \frac{\sum_{i \in V_c} N_i^{\text{intra}}}{\sum_{i \in V_c}(N_i^{\text{intra}} + N_i^{\text{inter}})}, \quad (5)$$

where V is a set of users in friendship network G. Using these measures, we investigated the strength of interaction locality in each dataset.

3.2 Macroscopic Analysis

We first perform macroscopic analysis, in which we investigate the strength of macroscopic locality using all communications data in the three datasets. **Table 2** shows the ratio of the number of intra-community interactions to the number of all interactions (i.e., the strength of macroscopic locality) in each dataset.

Table 2 shows that 80–90% of interactions are intra-community 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 communications in several types of media. This result offers evidence to prove the existence of strong interaction locality.

3.3 Microscopic Analysis

We next investigate the interaction locality at the microscopic level. We calculated the strength of microscopic locality for each user. **Figure 4** shows the distribution of the strengths of microscopic locality in each dataset. Note that users with no interactions are excluded in these figures.

Figure 4 shows that the strengths of the microscopic locality of most users are 0.8–0.9 in all datasets. This result shows that strong interaction locality can be observed not only from the macroscopic viewpoint, but also from the microscopic viewpoint. We can also find that small fractions of users do not have strong interaction locality. Such users might play a role to bridge multiple communities, or belong to multiple communities.

Table 3 Community size, number of intra-community calls (Intra calls), number of inter-community calls (Inter calls), and strength of mesoscopic locality by community (MIT dataset).

ID	Community size	Intra calls	Inter calls	Locality
1	13	1,991	55	0.97
2	13	1,418	14	0.99
3	9	2,393	184	0.93
4	9	690	18	0.97
5	4	288	159	0.64
6	3	20	0	1.00
7	3	10	29	0.26
8	2	8	0	1.00
9	2	0	5	0.00
10	2	4	15	0.21

3.4 Mesoscopic Analysis

We next investigate the interaction locality at the mesoscopic level. We investigate the strength of mesoscopic locality as well as other mesoscopic properties (i.e., community size, the number of intra-community interactions, and the number of inter-community interactions) for each community. Because the number of communities in the Facebook dataset is large (see Table 1) and showing the mesoscopic properties of all communities in the Facebook dataset would be lengthy, we focus on the MIT dataset and the Enron dataset in this subsection. The Facebook dataset will be investigated in the next subsection.

Tables 3 and **4** summarize several mesoscopic properties — community size, the number of intra-community interactions, the number of inter-community interactions, and the strength of mesoscopic locality — in the MIT dataset and the Enron dataset, respectively.

Table 3 shows that strong interaction locality is observed in most of the communities in the MIT dataset. Note that we should carefully check the strength of the mesoscopic locality in each community since some communities (in particular communities

Table 4 Community size, number of intra-community emails (Intra email), number of inter-community emails (Inter email), and strength of mesoscopic locality by community (Enron dataset).

ID	Community size	Intra email	Inter email	Locality
1	56	5,555	947	0.85
2	30	6,172	1,508	0.80
3	28	4,772	752	0.86
4	23	2,743	957	0.74
5	8	184	43	0.81
6	1	0	2	0
7	1	0	4	0
8	1	0	5	0
9	1	0	5	0

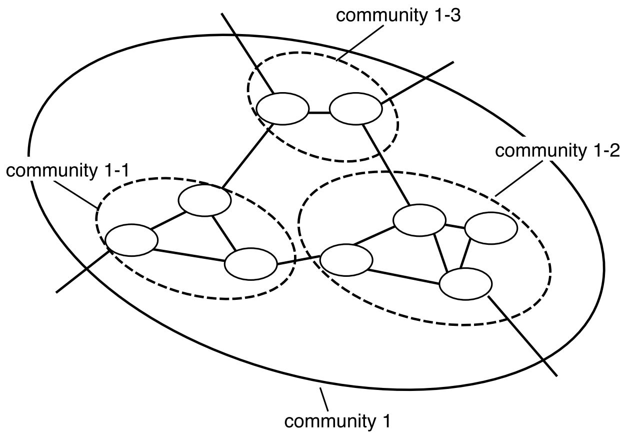


Fig. 5 An illustrative example of obtaining communities of different scales. Large community 1 is divided into smaller communities 1-1, 1-2, and 1-3.

8, 9, and 10) are very small. Table 4 also shows that strong interaction locality is observed in the Enron dataset.

We observe that the strengths of mesoscopic locality of different communities are comparable regardless of the sizes of communities. Moreover, we note that the strengths of macroscopic locality in the MIT dataset and the Enron dataset are also comparable. In other words, the strength of interaction locality of the Enron dataset as a whole is 0.82 (see Table 2), and the strengths of interaction locality of communities in the Enron dataset, which are obtained from partitioning, are still around 0.8 (see Table 4).

This result suggests a hypothesis that strong interaction locality is universal and independent from the scale of the community. Many social networks have hierarchical community structures [27]. Similar interaction locality strengths may be observed at different levels of the hierarchy (i.e., at different scales of communities). We further investigate this phenomenon in the next section using a large-scale dataset with hierarchical communities, the Facebook dataset.

3.5 Effect of the Scale of Community

To investigate the effects of the scale of community, we obtained communities at different scales and examined the strength of mesoscopic localities for these communities. An illustrative example of obtaining communities at different scales is shown in **Fig. 5**. To obtain communities at different scales, we recursively used the community detection algorithm [26]. We recursively partition a community in the Facebook dataset into smaller communities. Namely, we first obtain communities from the Facebook dataset. For each community, we apply the community detection algorithm to the subgraph corresponding to that community. More specifically, for subgraph $G_c = (V_c, E_c)$, where V_c is the set of nodes belonging to community c and E_c is the set of links connecting nodes in V_c, we divide G_c into communities. We then calculate the strength of mesoscopic locality in each obtained community. We repeat this procedure until the size of the community becomes smaller than or equal to 100.

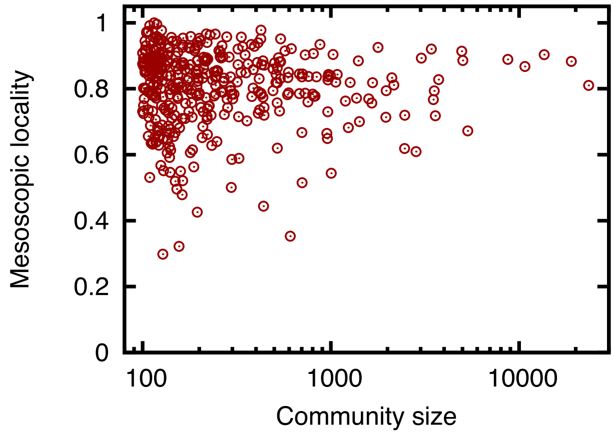


Fig. 6 Relation between the size of a community and the strength of its mesoscopic locality.

Figure 6 shows the relation between sizes and strengths of mesoscopic locality in all communities. This figure clearly shows the validity of our hypothesis; i.e., most communities have strong interaction locality, which is invariant to community sizes. The mean of the strengths of mesoscopic locality is 0.80. The standard error is 0.0059. We note that some smaller communities have weak interaction locality. The cause of this may be that community detection is too aggressive and thus artificially splits communities.

3.6 Cause of Scale Invariance

In this section, we discuss why the strength of the interaction locality is invariant with regard to the scale of a community. In the Facebook dataset, most (i.e., 97%) of the messages are exchanged between friends, so the strength of interaction locality is mostly determined by the ratio of the number of intra-community links to the number of all links in the friendship network. Therefore, a simple and intuitive explanation for scale-invariance in the interaction locality is that the community structure in the social relationship network itself is scale-invariant (i.e., the ratio of the number of intra-community links to the number of all links in a community is invariant regarding the scale of the community). Ferrara [28] shows that the ratio of intra-community links of most communities in Facebook are high regardless of the community size. We therefore expect that the community structure in Facebook could be scale-invariant.

We investigate the mesoscopic intra-community link ratio in each community (i.e., the ratio of the number of intra-community links to the number of all links in a community). **Figure 7** shows the relation between the size of a community and the intra-community link ratio in the community. This result shows that intra-community link ratios are centered around 0.8 regardless of the size of community. The mean of the intra-community link ratios is 0.79. The standard error is 0.0052. This result supports our hypothesis that community structure is scale-invariant. Therefore, we suggest that this is the main cause of the emergence of a scale-invariant pattern in interaction locality.

3.7 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 finally investigate the stationarity of the strength of interaction locality. We focus on the macroscopic locality rather than mesoscopic and microscopic locality to investigate the overall communication patterns. The time evolutions of the number of interactions and the strength of macroscopic locality per month are shown in **Figs. 8** and **9**, respectively. Time stamps of interactions are available in the datasets, and we calculated the macroscopic locality for each month from interactions occurred in the month.

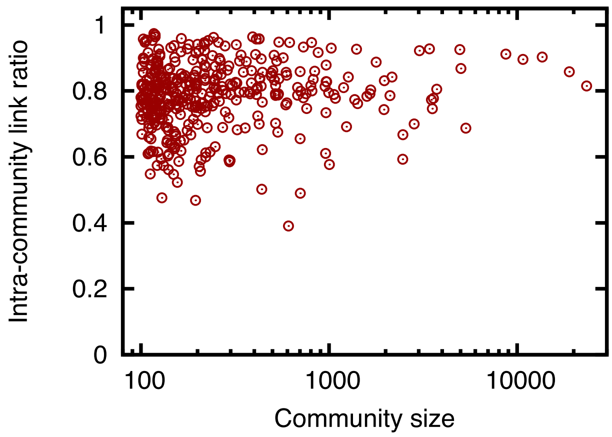


Fig. 7 Relation between the size of a community and the ratio of intra-community links to all links in the community.

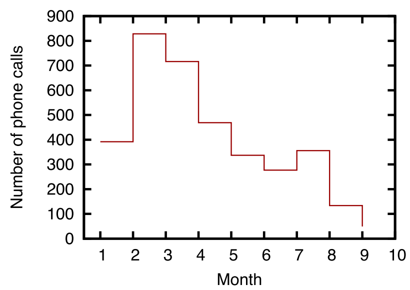
These figures indicate that in all datasets, the frequency of interactions significantly varies (see Fig. 8), but the strength of macroscopic locality is almost stable (see Fig. 9). Mean strengths of monthly macroscopic localities in MIT, Enron, and Facebook datasets are 0.94, 0.85, and 0.85, respectively. The standard errors are 0.0082, 0.016, and 0.0028, respectively. This suggests that the strength of interaction locality does not change frequently. A possible explanation on such finding is that neither the community structure in social relationship networks nor the communication patterns of the users change frequently.

4. Practical Implications

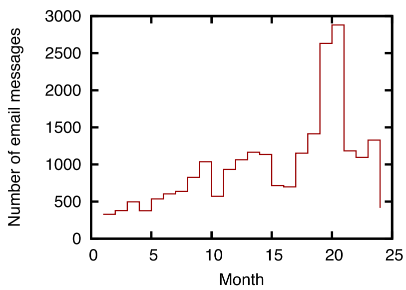
This section discusses practical implications of our results. The fact that strong interaction locality exists should be useful in designing new services.

For instance, our findings can be utilized for anomaly detection [29]. **Figure 10** illustrates anomalous communication detection using the interaction locality. Our results show the existence of strong interaction locality, which implies that a large volume of communication between members of different communities seems to be 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 suggests occurrence of possible anomalous events. Using these characteristics, a novel scheme for anomaly detection can be designed. Feasibility of anomaly detection based on our findings will be examined in Section 5.

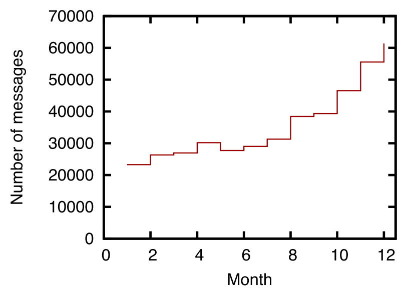
Another possible service based on our findings is estimation of traffic demands among individuals. Our results imply that coarse-grained traffic demands of individuals can be inferred from their communities. Such traffic pattern is expected to be useful for traffic engineering [30] and virtual network embedding [31]. **Fig-**



(a) MIT dataset

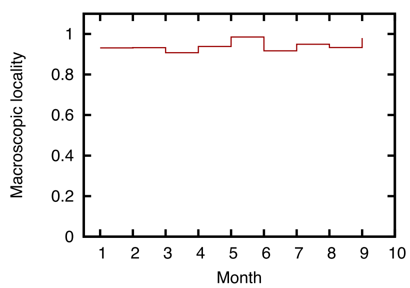


(b) Enron dataset

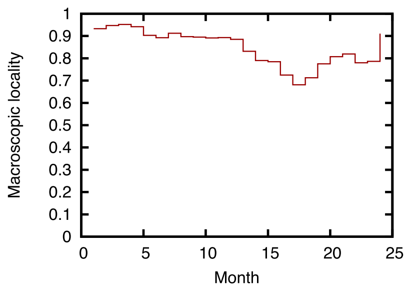


(c) Facebook dataset

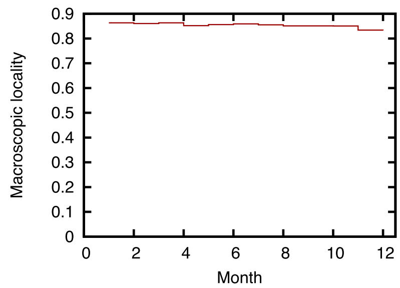
Fig. 8 Number of interactions for each month.



(a) MIT dataset



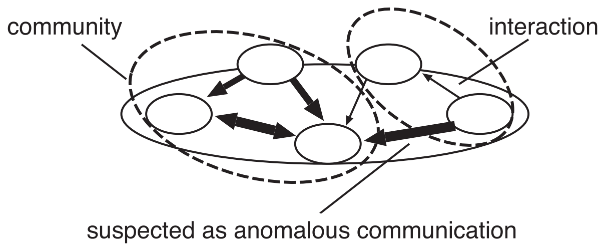
(b) Enron dataset



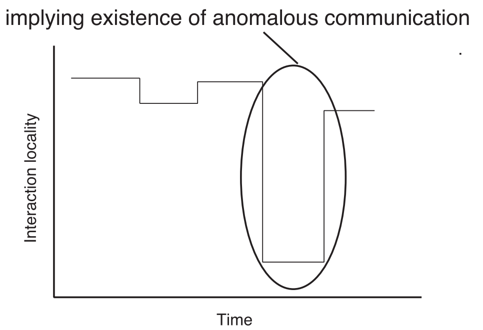
(c) Facebook dataset

Fig. 9 Strength of macroscopic locality for each month.

ure 11 illustrates an example of virtual network embedding using community information. Coarse-grained traffic demands can be estimated from network users' communities. We can construct a virtual network topology accommodating the estimated traffic demands. Since the traffic pattern is suggested to be stable, the constructed topology is expected to efficiently accommodate the traffic for long period of time. In our previous work [32], we have shown that efficient virtual network topology construction is achieved using community information if the strong interaction locality exists. When we conducted the previous work, there were little empirical evidence on the existence of the interaction locality, and the effectiveness of our method was not fully shown. However, our present work confirms the effectiveness of the previous work.



(a) A large amount of communication between members of different communities can be considered as anomalous



(b) Rapid changes in the strength of interaction locality can identify existence of anomalous communication

Fig. 10 Illustrative example of anomalous communication detection using the interaction locality.

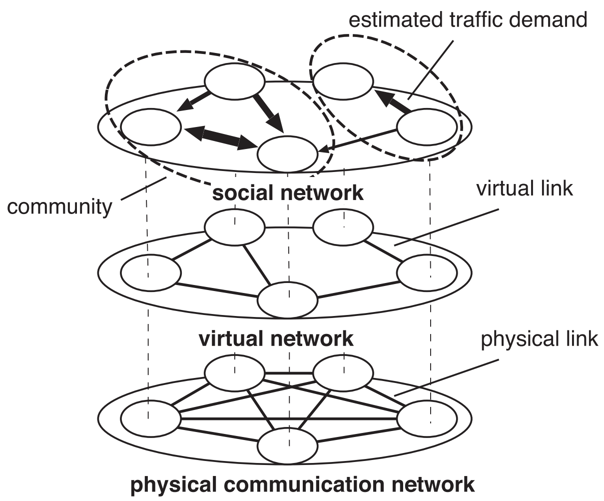


Fig. 11 An example of virtual network embedding using community information: A virtual network topology accommodating the estimated traffic demands obtained from community information is constructed [32].

The scale-invariance 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. Our results show that the strength of the interaction locality may be different for communication media. However, 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.

5. Case Study on Anomaly Detection based on Interaction Locality

Through a case study, we examine the feasibility of anomaly detection based on our results.

Anomaly detection is a problem to detect anomalous users who communicate anomalously. In what follows, we use the MIT dataset for the case study. We randomly select 10 users among 60 users in the MIT dataset, and synthetically make those users *anomalous* simply by multiplying the volume of their communications by $(1 + \alpha)$. Let N_i be a number of interactions (i.e., phone calls) from user i in the dataset. For each randomly-chosen anomalous user i, we randomly generate additional αN_i interactions from the user i to randomly-selected users.

For detecting anomalous users, our method utilizes knowledge obtained from our results. As shown in Section 3, the strength of microscopic locality of most users exceeds 0.9 (see Fig. 4 (a)). Our method calculates the strength of microscopic locality of a

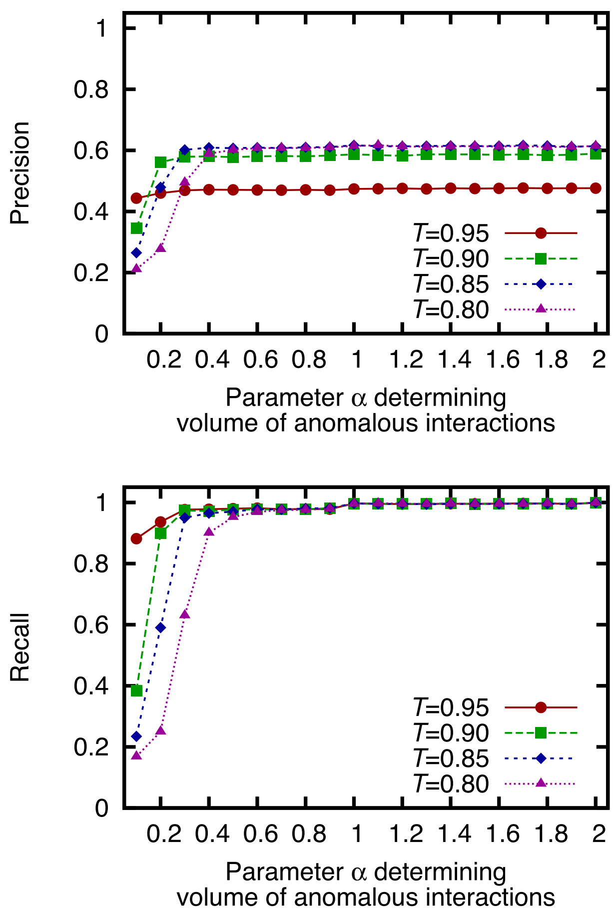


Fig. 12 Precision and recall when changing parameter α determining volume of anomalous interactions ($M = 0$).

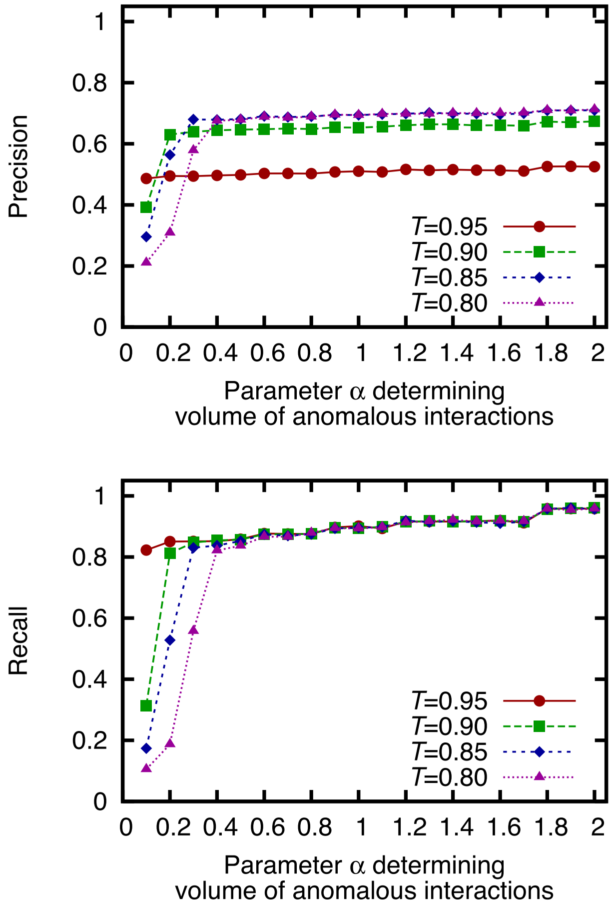


Fig. 13 Precision and recall when changing parameter α determining volume of anomalous interactions ($M = 10$).

user, and if it is less than a threshold T and the user has at least M interactions, the method classifies the user as anomalous.

We investigate how accurately those 10 anomalous users can be detected with our method while changing the parameter α. As measures for evaluating the accuracy of anomaly detection, we use precision and recall, which are standard measures for evaluating classification problems [33]. Let A be a set of anomalous users, and D be a set of users detected as anomalous with our method. Precision is defined as $|A \cap D|/|D|$, and recall is defined as $|A \cap D|/|A|$. We calculated average of these measures over 1,000 trials.

Figures 12 and **13** show precision and recall while changing the parameter α determining volume of anomalous communication when $M = 0$ and $M = 10$, respectively. Figures 12 and 13 show that if $\alpha > 0.3$, more than 80% anomalous users are detected with 60–70% precision by setting the threshold $T = 0.85$. These results show that our method successfully detect most anomalous users with considerable precision even when the volume of anomalous communication is small. This suggests usefulness of the interaction locality in anomaly detection, as well as the feasibility of anomaly detection using community information.

6. 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. 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 invariant with regard to the scale of the community. We also introduced possible applications of our results, and also showed the feasibility of anomaly detection using community information, which is one of the applications of our results.

As future work, we plan to design novel services based on the fact that strong interaction locality exists. For instance, as discussed in Section 4, anomalous interaction detection, traffic engineering, and virtual network embedding should be examples of possible applications. Preliminary experiment in Section 5 shows the potential of using the interaction locality in anomaly detection. However, extensive experiments using real anomaly data are required to fully investigate the effectiveness of the method. Study for constructing a model to estimate traffic demands of users is also considered to be an important future work.

Another direction of future work is further analysis of social relationship and interaction networks obtained from larger-scale and longer-period observations. There are some questions remaining unclear; e.g., (1) is the strength of interaction locality stable even if longer observation is performed?, and (2) are there any communities whose mesoscopic locality are extremely weak? To answer these and also to examine the generality of our findings in this paper, further investigations are necessary.

Acknowledgments This work was supported by JSPS KAKENHI Grant Number 25280030.

References

[1] Watts, D.J.: A twenty-first century science, *Nature*, Vol.445, No.7127, p.489 (2007).
[2] Borgatti, S.P., Mehra, A., Brass, D.J. and Labianca, G.: Network analysis in the social sciences, *Science*, Vol.323, No.5916, pp.892–895 (2009).
[3] Scott, J.: Social network analysis: Developments, advances, and prospects, *Social Network Analysis and Mining*, Vol.1, No.1, pp.21–26 (2011).
[4] Liu, X., Bollen, J., Nelson, M.L. and de Sompel, H.V.: Co-authorship networks in the digital library research community, *Information Processing & Management*, Vol.41, No.6, pp.1462–1480 (2005).
[5] Catanese, S.A., De Meo, P., Ferrara, E., Fiumara, G. and Provetti, A.: Crawling facebook for social network analysis purposes, *Proc. International Conference on Web Intelligence, Mining and Semantics* (*WIMS'11*), p.52 (2011).
[6] Mislove, A., Koppula, H.S., Gummadi, K.P., Druschel, P. and Bhattacharjee, B.: Growth of the Flickr Social Network, *Proc. 1st Workshop on Online Social Networks* (*WOSN'08*), pp.25–30 (2008).
[7] Ahn, Y.-Y., Han, S., Kwak, H., Moon, S. and Jeong, H.: Analysis of topological characteristics of huge online social networking services, *Proc. 16th International Conference on World Wide Web* (*WWW'07*), pp.835–844 (2007).
[8] Huang, Z. and Zeng, D.D.: Signed Networks in Social Media, *Proc. SIGCHI Conference on Human Factors in Computing Systems* (*CHI'10*), pp.1361–1370 (2010).
[9] Holger, E., Lutz-Ingo, M. and Stefan, B.: Scale-free topology of e-mail networks, *Physical Review E* (*Statistical, Nonlinear, and Soft Matter Physics*), Vol.66, No.3, p.035103 (2002).
[10] Kossinets, G. and Watts, D.J.: Empirical analysis of an evolving social network, *Science*, Vol.311, No.5757, pp.88–90 (2006).
[11] Onnela, J.P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J. and Barabási, A.L.: Structure and tie strengths in mobile communication networks, *Proc. National Academy of Sciences*, Vol.104, No.18, pp.7332–7336 (2007).
[12] Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.-F. and den Broeck, W.V.: What's in a crowd? Analysis of face-to-face behavioral networks, *Journal of Theoretical Biology*, Vol.271, No.1, pp.166–180 (2011).

[13] Ehrlich, K. and Cataldo, M.: All-for-one and One-for-all?: A Multi-level Analysis of Communication Patterns and Individual Performance in Geographically Distributed Software Development, *Proc. ACM 2012 Conference on Computer Supported Cooperative Work* (*CSCW'12*), pp.945–954 (2012).
[14] Cataldo, M. and Ehrlich, K.: The impact of communication structure on new product development outcomes, *Proc. SIGCHI Conference on Human Factors in Computing Systems* (*CHI'12*), pp.3081–3090 (2012).
[15] Newman, M.E.J. and Girvan, M.: Finding and evaluating community structure in networks, *Physical Review E*, Vol.69, No.2, p.026113 (2004).
[16] Lomi, A. and Larsen, E.R.: Interacting locally and evolving globally: A computational approach to the dynamics of organizational populations, *Academy of Management Journal*, Vol.39, No.5, pp.1287–1321 (1996).
[17] Leifer, E.M.: Interaction preludes to role setting: Exploratory local action, *American Sociological Review*, Vol.53, No.6, pp.865–878 (1988).
[18] Mandel, M.J.: Local roles and social networks, *American Sociological Review*, Vol.48, No.3, pp.376–386 (1983).
[19] Tsugawa, S., Ohsaki, H., Itoh, Y., Ono, N., Kagawa, K. and Takashima, K.: Dynamic Social Network Analysis with Heterogeneous Sensors in Ambient Environment, *Social Networking*, Vol.3, No.1, pp.9–18 (2014).
[20] Eagle, N., Pentland, A. and Lazer, D.: Inferring friendship network structure by using mobile phone data, *Proc. National Academy of Sciences*, Vol.106, No.36, pp.15274–15278 (2009).
[21] Shetty, J. and Adibi, J.: The Enron email dataset database schema and brief statistical report, Technical Report, Information Sciences Institute, University of Southern California (2004).
[22] Viswanath, B., Mislove, A., Cha, M. and Gummadi, K.P.: On the Evolution of User Interaction in Facebook, *Proc. 2nd ACM SIGCOMM Workshop on Social Networks* (*WOSN'09*), pp.37–42 (2009).
[23] Golder, S., Wilkinson, D. and Huberman, B.: Rhythms of social interaction: Messaging within a massive online network, *Proc. 3rd International Conference on Communities and Technologies* (*CT 2007*), pp.41–66 (2007).
[24] Weng, L., Ratkiewicz, J., Perra, N., Gonçalves, B., Castillo, C., Bonchi, F., Schifanella, R., Menczer, F. and Flammini, A.: The Role of Information Diffusion in the Evolution of Social Networks, *Proc. 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD'13*), pp.356–364 (2013).
[25] Palla, G., Barabási, A.-L. and Vicsek, T.: Quantifying social group evolution, *Nature*, Vol.446, No.7136, pp.664–667 (2007).
[26] Clauset, A., Newman, M.E.J. and Moore, C.: Finding community structure in very large networks, *Physical Review E*, Vol.70, No.6, p.066111 (2004).
[27] Clauset, A., Moore, C. and Newman, M.E.J.: Hierarchical structure and the prediction of missing links in networks, *Nature*, Vol.453, No.7191, pp.98–101 (2008).
[28] Ferrara, E.: A large-scale community structure analysis in Facebook, *EPJ Data Science*, Vol.1, No.1 (2012).
[29] Huang, Z. and Zeng, D.D.: A link prediction approach to anomalous email detection, *Proc. 2006 IEEE International Conference on Systems, Man, and Cybernetics* (*SMC'06*), pp.1131–1136 (2006).
[30] Elwalid, A., Jin, C., Low, S. and Widjaja, I.: MATE: MPLS adaptive traffic engineering, *Proc. 20th Annual Joint Conference of the IEEE Computer and Communications Societies* (*INFOCOM'01*), pp.1300–1309 (2001).
[31] Leivadeas, A., Papagianni, C. and Papavassiliou, S.: Socio-aware virtual network embedding, *IEEE Network*, Vol.26, No.5, pp.35–43 (2012).
[32] Ogasa, T., Takahashi, Y., Ohsaki, H., Imase, M., Yagi, T. and Murayama, J.: Dynamic topology reconfiguration method for service overlay networks using users' community information, *Proc. 9th Annual International Symposium on Applications and the Internet* (*SAINT'09*), pp.141–144 (2009).
[33] Yang, Y.: An evaluation of statistical approaches to text categorization, *Information retrieval*, Vol.1, No.1-2, pp.69–90 (1999).

Sho Tsugawa received his M.E. and Ph.D. degrees from Osaka University, Japan, in 2009 and 2012, respectively. He is currently an assistant professor at the Faculty of Engineering, Information and Systems, University of Tsukuba, Japan. His research interests include social network analysis in large-scale online communities. He is a member of IEEE, IEICE, and IPSJ.

Hiroyuki Ohsaki received his M.E. degree in the Information and Computer Sciences from Osaka University, Japan, in 1995. He also received Ph.D. degree from Osaka University, Japan, in 1997. He is currently a professor at Department of Informatics, School of Science and Technology, Kwansei Gakuin University, Japan. His research work is in the area of design, modeling, and control of large-scale communication networks. He is a member of IEEE, IEICE, and IPSJ.