In: Proceedings of the LWA - Learning, Knowledge, Adaptation.
University of Bamberg, Germany. 2013

Evolution of Contacts and Communities in Networks of Face-to-Face Proximity

Extended Abstract*

Mark Kibanov Martin Atzmueller Christoph Scholz Gerd Stumme
Knowledge and Data Engineering Group, University of Kassel
Wilhelmshöher Allee 73, D-34121 Kassel, Germany
Email: {kibanov,atzmueller,scholz,stumme}@cs.uni-kassel.de

Abstract

Communities are a central aspect in the formation of social interaction networks. In this paper, we analyze the evolution of communities in networks of face-to-face proximity. As our application context, we consider four scientific conferences. We compare the basic properties of the contact graphs to describe the properties of the contact networks and analyze the resulting community structure using state-of-the-art automic community detection algorithms. Specifically, we analyze the evolution of contacts and communities over time to consider the stability of the respective communities. In addition, we assess different factors which have an influence on the quality of community prediction. Overall, we provide first important insights into the evolution of contacts and communities in face-to-face contact networks.

1 Introduction

In this paper, we consider the evolution of both contacts and communities at academic conferences. Specifically, we consider the LWA 2010, LWA 2011, LWA 2012 and Hypertext 2011 conferences, where the CONFERATOR[1] system [1] was applied. Using RFID technology, it allows us to collect face-to-face contact data [3], which we can utilize for analyzing contacts and communities.

Our contribution is summarized as follows:

1. We analyze if the structure of the contact graphs is similar for different conferences.
2. We investigate the progress of face-to-face contacts during the respective conferences.
3. We consider automatically detected communities, and analyze the quality of the used algorithms.
4. Finally, we analyze how communities develop over time during a conference and whether detected communities stay stable and thus predictable.

To the best of the authors' knowledge, this is the first time, that these research questions have been addressed in the context of human face-to-face contact networks.

*This extended abstract summarizes the paper [4]: Mark Kibanov, Martin Atzmueller, Christoph Scholz, and Gerd Stumme. On the Evolution of Contacts and Communities in Networks of Face-to-Face Proximity. Proc. IEEE CPSCom 2013, IEEE Computer Society, Boston, MA, USA, 2013

[1] `http://www.conferator.org`

2 Analysis

In the following, we first briefly describe the utilized dataset, before we summarize the evolution of contacts and communities. For a detailed discussion, we refer to [4].

2.1 Datasets

At the LWA 2010, 2011, 2012 and Hypertext 2011 conferences we asked each participant to wear proximity tags, so they could use the CONFERATOR [1] system. These tags can detect close-range face-to-face proximity (1-1.5 meters) of the participants wearing them [3] - 77 (LWA 2010), 69 (Hypertext), and 42 (LWA 2011 and LWA 2012) for the respective conferences.

2.2 Evolution of Contacts

In summary, the number of edges in contact graph grows nearly linearly during all three LWA conferences. The number of new contacts at the beginning and at the end of these conferences can be explained by the small number of participants who come early or stay longer. An interesting fact for the Hypertext conference is a slow growth of contacts during the second part of the conference. This "tail" is much longer compared to the end of the LWA conferences. We assume that the Hypertext conference has a different "social profile", so the participants are more focused on "socializing" during the first day.

Another important observation shows that graphs with "long" talks ($\geq$ 180 seconds) have almost half of the number of edges of the graphs with all conversations, but their total length is equal to 80% – 90% of the whole length of the whole graph.

2.3 Evolution of Communities

For analyzing the stability of community structure we define a *c-pair* (*Community-pair*) as follows: If two nodes u and v belong to the same community, then $cp = (u, v)$ is a c-pair. CP denotes the set of all possible *c-pairs*. The more c-pairs stay over time, the more stable is a community structure.

To estimate and compare the stability of communities during different conferences, we applied a "simple" predictor $P : I \times J \rightarrow CP$, where $I \subseteq \mathbb{N}, J \subseteq \mathbb{N}$. This predictor assumes that all the c-pairs that were built during (a) reference day(s) in I will be also formed during the subsequent day(s) in J. In the case where I and J contain only single elements, we will drop the set notation for simplicity. Let CP_i be the set of c-pairs of day i: $CP_i = \{(u, v) \mid u, v \in C_j \subseteq V_i\}$, where V_i is the set of the nodes of the contact graph of the day i. We applied

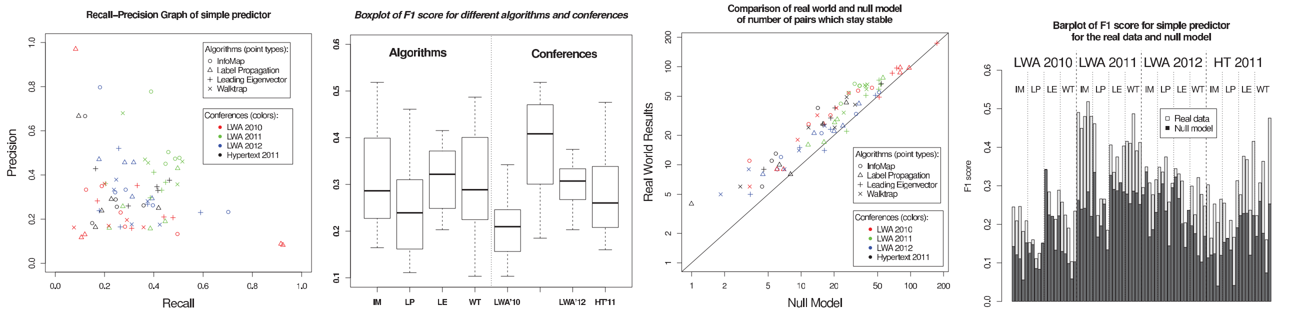

Figure 1: (a) Recall-Precision Graph of the "simple" predictor of the considered algorithms (marked by point types) and conferences (marked by colors). (b) Boxplots of the F1 score of different algorithms (on the left side) and conferences (on the right side). The abbreviations mean: IM – InfoMap, LP – Label Propagation, LE – Leading Eigenvector, WT – Walktrap. (c) Comparison of the real community stability with the null model of the considered algorithms (marked by the different point types) and conferences (marked by different colors): The x-axis contains the respective null model values, the y-axis contains the respective real values. Both axes are scaled logarithmically. (d) Barplots of the F1 scores of the predictions compared to the respective null model for different conferences and different algorithms shown on the top.

the predictor five times for each algorithm and each conference. For computing the 'correct' predictions, we consider the intersection with a subsequent day, and the respective c-pairs. The more c-pairs are predicted correctly, the more stable is the computed community structure.

Figure 1 shows the respective recall and precision values. The larger the value of precision, the more c-pairs from the "training"-day tend to appear also during the "result"-day. The larger the value of recall, the less new c-pairs tend to appear during the "result" day. The type of the point defines the applied algorithm and the color of the point defines the conference. Red circles, for example, show recall and precision of predictions made by the InfoMap algorithm for the LWA 2010 conference. The LWA 2011 data (green points) tend to show a better performance compared to the other conferences and thus we assume the community structure during LWA 2011 is more stable. Similarly, the communities of LWA 2012 are also rather well "predictable". A potential explanation is given by the significant community structure of the four special interest groups constituting the LWA conferences, see [2]. Summarizing both precision and recall, the F1 scores for each applied algorithm and each conference are shown in Figure 1.

The choice of the community detection algorithm did not have a big impact on the performance of our "simple" algorithm and thus on the obtained communities. On the other hand, the choice of the event has a crucial influence on the stability of the communities: The F1 scores confirm the stability of community structure computed for the LWA 2011 conference (green points). The stability of the community structures detected for the LWA 2012 conference show the smallest deviation (The F1 score lies between 0.2 and 0.4).

As another interesting observation, the active communication does not make communities stable – even vice versa. Comparing the LWA 2011 and LWA 2012 conferences with the similar number of participants, we see that the LWA 2012 communications were less active than those at the LWA 2011 in terms of graph density and the total length of communication; overall, we observe more stable communities during LWA 2011. We observed the same phenomenon considering LWA 2010 and HT 2011 – two conferences with the same number of participants but very different dynamics of face-to-face communications. On hypothesis for explaining the negative correlation of community stability and communication is the following: The participants stick to the known persons and tend to have less contacts with new persons which implies both lack of new contacts and stability of the existing communities over the whole conference.

So far, our proposed measures compare the overall stability of communities of different conferences. However, in order to clarify that these stabilities are significant and not accidental, we apply a null model NM computed using the following formula: $NM = 2 \times \frac{CP_t}{n \times (n-1)} \times CP_{t+1}$, where CP_i is the number of c-pairs at day i, and n is the number of nodes in the considered graph. As shown in Figure 1, the majority of points lies above the null model line which means the stability of communities is not a random phenomenon. Some of the results obtained using the $LeadingEigenvector$ algorithm lie below the null model line, while some of the $LabelPropagation$ measurements are just placed on the line. These findings would seem to show some randomness of the stability of community structures computed with these algorithms. In order to characterize the stability further, we compare the F1 score of the real data and the null model (see Figure 1). On average the real world F1 score is 1.65 times larger than the obtained null model F1 score. This shows, that persons tend to stay in the same communities over one conference; the choice of algorithm also does not affect this.

References

[1] M. Atzmueller, D. Benz, S. Doerfel, A. Hotho, R. Jäschke, B. E. Macek, F. Mitzlaff, C. Scholz, and G. Stumme. Enhancing Social Interactions at Conferences. *it*, 53(3):101–107, 2011.

[2] M. Atzmueller, S. Doerfel, A. Hotho, F. Mitzlaff, and G. Stumme. Face-to-Face Contacts at a Conference: Dynamics of Communities and Roles. In *Modeling and Mining Ubiquitous Social Media*, volume 7472 of *LNAI*. Springer Verlag, Heidelberg, Germany, 2012.

[3] A. Barrat, C. Cattuto, V. Colizza, J.-F. Pinton, W. V. den Broeck, and A. Vespignani. High Resolution Dynamical Mapping of Social Interactions with Active RFID. *PLoS ONE*, 5(7), 2010.

[4] M. Kibanov, M. Atzmueller, C. Scholz, and G. Stumme. On the Evolution of Contacts and Communities in Networks of Face-to-Face Proximity. In *Proc. IEEE CPSCom 2013*, Boston, MA, USA, 2013.