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

Stefan Bloemheuvel, Martin Atzmueller, and Marie Postma

Tilburg University, The Netherlands

{s.d.bloemheuvel,m.atzmuller,marie.postma}@tilburguniversity.edu

Abstract. Temporal evolution and dynamics of social network interactions provide insights into the formation of social relationships. In this paper, we explore automatic detection of face-to-face proximity during two student meet-ups for the purposes of community detection. The data was collected with the help of wearable sensors. Next to community detection, we examine the structural metrics of the formed networks over time. Contrary to previous studies, we observed that conversations tended to develop in a parabolic rather than linear manner during both events. For community detection, overall the *Louvain* method showed the most promising consistent results for both networks.

1 Introduction

Social relationships that are formed during events can be captured in face-to-face contact networks [1, 3]. Their analysis can provide insights into the dynamics, predictability [30–32] and evolution of communities, e. g., cf. [4, 19, 20]. In the past, the available methods to collect empirical data relied on surveys and diary methodologies which are slow and inaccurate [22]. Therefore, novel technologies have been developed which provide new and promising approaches of collecting face-to-face contact data [2, 6, 14, 38].

In this paper, we focus on face-to-face proximity as the basic measure of social contact occurrence and investigate the evolution of contacts and communities in social networks. For data collection, we utilized wearable sensors developed by the SocioPatterns consortium.[1] The proximity tags are based on Radio Frequency Identification technology (RFID chips), capable of detecting close-range and face-to-face proximity (1 - 1.5 meters) with a temporal resolution of 20 seconds [10]. The most significant advantage of using the SocioPatterns tags is that the human body acts as a radio-frequency blocker at the frequency used on the chips [10]. Therefore, only the signals that are broadcasted directly forward from the person wearing the tag will be detected by other tags. With help of our own bodies, a face-to-face contact can be observed with a probability of over 99 percent using the interval of 20 seconds as the minimal contact duration [10].

The goals of this work were twofold: First to assess the evolution of contacts of face-to-face proximity, and second to compare the performance of existing

[1] http://www.sociopatterns.org

Preprint of: Stefan Bloemheuvel, Martin Atzmueller, and Marie Postma (2018)
Evolution of Contacts and Communities in Social Interaction Networks of Face-to-Face Proximity.
In: Proc. Benelux Conference on Artificial Intelligence (BNAIC), Jheronimus Academy of Data Science (JADS), 's-Hertogenbosch, The Netherlands

algorithms for community detection, i. e., for detecting "densely connected" areas of a network. Practical applications of such algorithms include monitoring human activity to study the dynamics of human contacts, such as hospitals, museums and conferences [4, 18, 34]. Community detection is rather challenging in that a network can be divisible into several communities according to various criteria, sometimes also with a hierarchical structure [15, 24]. The availability of competent community detection algorithms is therefore of utmost importance. For the automatic detection of communities, several algorithms can be used, e. g., [15, 24], for which standard methods are implemented in the igraph [12] software package. The general problem definition of community detection can be formulated as follows: given a network $G(V, E)$, find the optimal communities as in closely connected groups of nodes and a moderate number of disparate outliers [35]. We applied the following algorithms from recently proposed taxonomies of numerous community detection methods [11, 15, 37]: Edge Betweenness, Info Map, Spinglass, Louvain, Label Propagation, and Leading Eigenvector.

All of these algorithms differ substantially. Edge Betweenness is a hierarchical process decomposing a given graph [24] by removing edges in decreasing order of their edge betweenness scores, i. e., identifying edges on multiple shortest paths which in many cases are bridges linking groups together. Infomap is based on random walks and information theoretic principles [29]. The algorithm tries to group based on the shortest description length for a random walk on the graph. The description length is measured by the number of bits per vertex required to encode the path of the random walk. The best communities are those that transmit a large deal of information while requiring minimal bandwidth. Spinglass originates from physics and is based on the Potts model [28]. Each node can be in a certain state and the interactions between the nodes (the edges) specify which state the node has. This process is simulated several times and nodes with the same state are seen as a community. Spinglass provides a parameter determining the cluster sizes, i. e., thus the number of communities. The Louvain method is a top-down process based on modularity optimization [7]. First, small communities are identified by their modularity score [17, 24] (each node is its own community); smaller communities are then grouped together if and only if it increases the modularity. This process is repeated until the merging of communities will no longer increase the modularity score. Label Propagation is the simplest approach in which every node is given one of k labels [27]. During each iteration, every label a node has is re-assigned in a manner that nodes adopt the most frequent label of its neighbors. The algorithm stops when every node has the most frequent label in its neighborhood. Finally, Leading Eigenvector is modularity based and uses optimization, inspired by a technique called graph partitioning [23]. The algorithm tries to find the eigenvector that corresponds to the most positive eigenvalue of the modularity matrix. Afterwards, it divides the network into communities in harmony with the elements of the vector.

In the rest of the paper, we provide information describing the data collection and preprocessing for the purposes of network detection analysis in Section 2, the results of the analysis using the six algorithms described above in Section 3, and some insights regarding the nature of the communities detected in Section 4.

2 Data Set

The data was collected during two student events organized at a university in the Netherlands. Attendees were invited to participate by signing a consent form. The data collection was anonymized, including solely information about participant gender. For the two events, the number of male and female participants was comparable (event 1: 12 male and 11 female; event 2: 10 male and 9 female). The age interval was between 18 to 24 years old. Each participant was asked to wear a SocioPatterns proximity tag. Social contact was established when the contact between the proximity tags was at least 20 seconds [10]. Interactions that took place within an interval of 20 seconds were merged if the actors remained the same, according to the commonly used threshold [10]. This resulted in a data set where accidental interactions were filtered out (e. g., two people crossing each other a total of 4 seconds to a different area of the event). In addition, a minimum RSSI value of -85 was used as a threshold to filter out weak interactions.

3 Results

As shown in Table 1, a total of 25 participants used the proximity tags at event 1 and 19 at event 2. Comparing both graphs, some structural differences can be observed w.r.t. contact duration and the number of interactions (see Table 1). A possible explanation of this difference might be the nature of the event; event 1 was organized between two groups that did not knew each other well. The network of event 2 shows a diameter of 2, which indicates a small-world effect [9].

Table 1: High level network statistics: Number of nodes (N) and edges (E), Average contact duration (Avg. Con), Longest contact (L), Network diameter (D), Graph density (Density), Transitivity and Average path length (Avg. Path).

Network	N	E	Avg. Con	L	D	Density	Transitivity	Avg. Path
Event 1	25	606	309.60	5807	5	2.02	0.47	2
Event 2	19	1239	535.70	5126	2	7.25	.60	1.51

Table 2 shows several centrality measures [8, 16]: the individual measures show which participants played an important role during the events.

Table 2: Average node centrality measures for both events.

Network	Eigenvector	Average Degree	Closeness	Betweenness
Event 1	0.21	48.48	0.02	12
Event 2	0.29	130.42	0.04	4.58

Event 2 showed a higher eigenvector centrality than event 1. A high eigenvector score suggests that all nodes are connected to other nodes that have a high eigenvector score, which indicates that the network of event 1 was densely interconnected. The degree centrality of event 2 was higher than event 1, indicating that on average, the participants during event 2 had more connections (proportionally, since the networks differ in the number of participants). Both events show a low closeness score of 0.02-0.04 indicating that each participant could be reached in relatively few steps. In Figure 1, the aggregated contact duration of both events is shown. In line with a commonly seen power law [6, 10, 21], the shorter a contact was, the more likely it was to occur.

In order to analyze the evolution of contacts, the events were divided into three intervals. In Table 3 and 4, the evolution of several measures of the network of event 1 and 2 is shown. During each period of event, almost all nodes were active (see Table 3). A different distribution is visible when examining the number of edges that are present in each period. The number of active edges decreases in event 1 in period 2, but recovers during period 3. The diameter scores confirm this observation. The diameter was lowest during period 2 of event 1 (indicating a small world effect) and highest during period 1 and 3. A similar pattern as the evolution of edges and diameter is visible in the e dge density and average degree. Table 4 shows the network evolution metrics of event 2. The average contact duration decreases between period 1 to period 2 and rises again in period 3. Similar to event 1, not all nodes were active during each period of the event and the number of edges increases in the second and even the third period. Finally, in Figure 2 and 3, the changes in the Betweenness centrality are clearly represented.

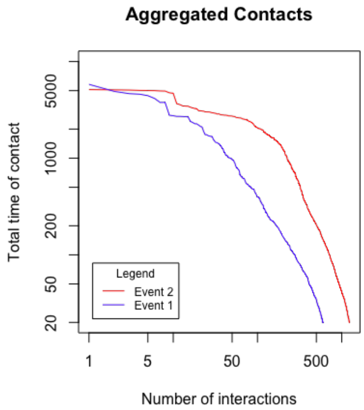


Fig. 1: Cumulated contact length (in seconds) distribution of all face-to-face contacts.

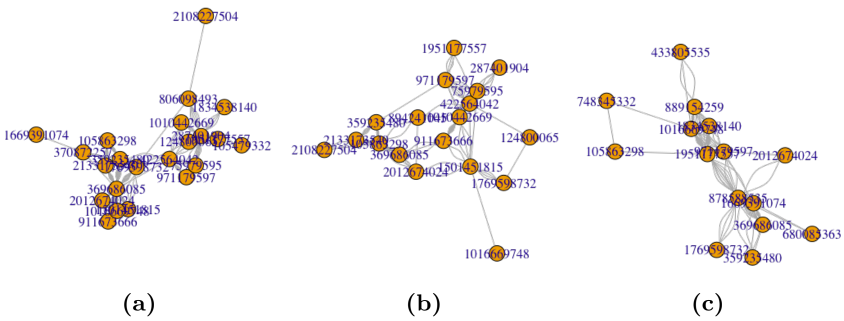


Fig. 2: Event 1 divided into three equal slices. From left to right: (a) period 1, (b) period 2 and (c) period 3.

Table 3: Graph evolution metrics of Event 1: Number of nodes (N), edges (E), Diameter, Average path length, Longest contact length (in seconds), Average contact length (in seconds), Edge density, Average Degree, Closeness and Betweenness centrality.

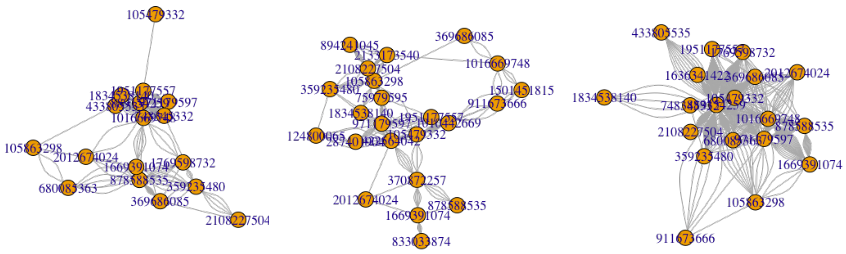
	Period 1	Period 2	Period 3
N	23	18	22
E	268	145	193
Diameter	5	4	5
Avg. Path Length	2.41	2.24	2.50
Max Contact Length	5807	2783	1009
Avg. Contact Length	422.97	369.10	107.47
Edge Density	1.06	0.95	0.84
Avg. Degree	23.30	16.11	17.55
Avg. Closeness	0.02	0.03	0.02
Avg. Betweenness	15.5	10.50	15.77

Table 4: Graph evolution metrics of Event 2: Number of nodes (N), edges (E), Diameter, Average path length, Longest contact length (in seconds), Average contact length (in seconds), Edge density, Average Degree, Closeness and Betweenness centrality.

	Period 1	Period 2	Period 3
N	15	17	19
E	150	411	678
Diameter	3	5	3
Avg. Path Length	1.94	2.01	1.64
Max Contact Length	5126	3445	1630
Avg. Contact Length	1041.07	916.58	193.02
Edge Density	1.43	3.02	3.95
Avg. Degree	20	48.35	71.37
Avg. Closeness	0.04	0.03	0.03
Avg. Betweenness	6.6	8.06	5.79

It is important to note that the results of period 3 in event 2 are negatively influenced by the fact that the graph becomes sparsely connected by a two bridges. Especially the closeness centrality suffers from a disconnected graph since the metric calculates the distance between nodes. If the network is disconnected, the distance is infinite [25, 36]. One might not notice this effect in Table 4, but the closeness centrality decline from 0.02 to 0.04 was large since the values of the closeness centrality metric tend to span a moderately small dynamic range [25].

Since we did not collect any information about individual participants apart from their gender, we were unable to analyze the effect of sociodemographic variables on community emergence. With respect to gender, we observed clear signs of homophily. During event 1, both groups barely knew each other before the start of the event. During event 2, both groups have had several earlier events



(a) (b) (c)

Fig. 3: Event 2 divided into three equal periods. From left to right: (a) period 1, (b) period 2 and (c) period 3.

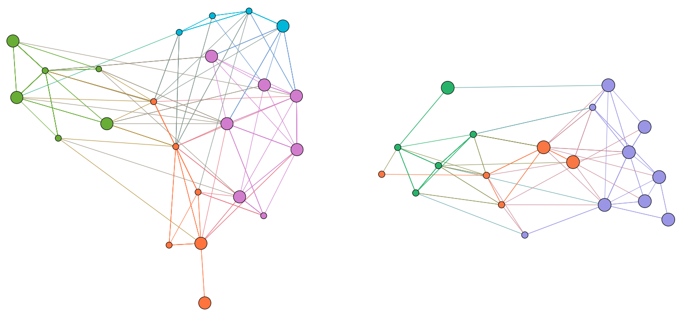
together. In Figure 4, the networks of both events are depicted, where exemplary communities are shown – detected using the Louvain algorithm.

In order to analyze the robustness of the community detection algorithms, we used the modularity score of the total graph and the average computational time (each algorithm was run a total of 20 times and the time was averaged). On average, the modularity score, which gives insight in the strength of the divisions of a network into communities, was 0.27 (SD=0.19) for event 1 and 0.08 (SD=0.07) for event 2. No algorithm managed to determine communities with a modularity higher than 0.46 (see Table 5). The Spinglass algorithm performed worst on both events and took the longest to compute. The Louvain and Label Propagation algorithms showed the highest modularity scores for event 1 and achieved this result in a short computational time.

The community structure for an event can be examined using the normalized mutual information metric (NMI) [13]. The more basic variant mutual information (MI) is a measure of the mutual dependence between two variables [13]. Since NMI is normalized, we can measure and compare the NMI score of the resulted networks in regard to their calculated communities (NMI = 0 means no mutual information, NMI = 1 means perfect correlation). Table 6 and 7 show the NMI scores of each algorithm for event 1 and 2.

4 Conclusion and Discussion

We employed RFID chips with proximity detection to collect face-to-face contact data during two events. We performed an analysis of the contact graphs and examined the evolution of contacts by dividing the contact networks into three different periods. In order to automatically detect communities, six conceptually different community detection algorithms were used. To analyze the quality of the communities that were detected, their modularity score, the number of detected communities, the computational time each algorithm needed and normalized mutual information were examined. Conversations tended to develop in a parabolic manner during both events. This parabolic tendency is contradictory



(a) **(b)**

Fig. 4: Communities detected using Louvain (Modularity) where male nodes are sized big and female nodes are sized small: (a) event 1, (b) event 2.

Table 5: Comparison of several community detection algorithms for both events. The communities, modularity and computational time are the averages of 20 simulations

Algorithm	Parameter	Event 1	Event 2
Edge Betweenness	No. of Communitites	6	8
	Modularity	0.24	0.03
	Computational Time	3.03	12.07
Info Map	No. of Communitites	10	7
	Modularity	0.04	0.09
	Computational Time	0.37	0.41
Spinglass	No. of Communitites	4	3
	Modularity	0.07	0.03
	Computational Time	11.54	10.93
Louvain	No. of Communitites	4	3
	Modularity	0.46	0.17
	Computational Time	0.07	0.07
Label Prop	No. of Communitites	4	1
	Modularity	0.46	0.00
	Computational Time	0.05	0.06
Leading Eigenvector	No. of Communitites	5	3
	Modularity	0.39	0.15
	Computational Time	1.44	1.16

Table 6: Comparison of community structures using NMI on event 1.

Comparisons								
Edge Betweenness	Info Map	0.58	Info Map	Edge Betweenness	0.58	Spinglass	Edge Betweenness	0.86
	Spinglas	0.86		Spinglas	0.53		Info Map	0.53
	Louvain	0.92		Louvain	0.56		Louvain	0.83
	Label Prop	0.79		Label Prop	0.51		Label Prop	0.68
	Leading Eig	0.80		Leading Eig	0.61		Leading Eig	0.73
Louvain	Edge Betweenness	0.92	Label Prop	Leading Eig	0.72	Leading Eig	Edge Betweenness	0.80
	Info Map	0.56		Edge Betweenness	0.79		Info Map	0.61
	Spinglas	0.83		Info Map	0.51		Spinglass	0.73
	Label Prop	0.87		Spinglass	0.68		Louvain	0.86
	Leading Eig	0.86		Louvain	0.87		Label Prop	0.72

Table 7: Comparison of community structures using NMI on event 2.

Comparisons								
Edge Betweenness	Info Map	0.53	Info Map	Edge Betweenness	0.53	Spinglass	Edge Betweenness	0.55
	Spinglas	0.55		Spinglas	0.37		Info Map	0.37
	Louvain	0.51		Louvain	0.24		Louvain	0.20
	Label Prop	0.00		Label Prop	0.00		Label Prop	0.00
	Leading Eig	0.56		Leading Eig	0.33		Leading Eig	0.35
Louvain	Edge Betweenness	0.52	Label Prop	Leading Eig	0.00	Leading Eig	Edge Betweenness	0.56
	Info Map	0.24		Edge Betweenness	0.00		Info Map	0.33
	Spinglas	0.20		Info Map	0.00		Spinglass	0.35
	Label Prop	0.00		Spinglass	0.00		Louvain	0.51
	Leading Eig	0.51		Louvain	0.00		Label Prop	0.00

to the linear patterns other studies have found [19, 20]. The most likely explanation is that the linear patterns were found for events with a highly structured program (for example, the end of the presentation of the key speaker would likely result in a peak of interactions at the end of the event).

With respect to community detection, the Label Propagation and Louvain algorithms showed the most promising results in the network of event 1. Both their modularity scores ranked the highest while demanding the least computational time. The lack of results for the Leading Eigenvector and Label Propagation algorithms is in line with earlier studies [26]. It is interesting to note that the algorithm performs poorly in both an experimental (in most studies the networks are randomly generated) and the real-life setting.

Most noteworthy, the Label Propagation algorithm failed to calculate any valuable information in the network of event 2. It is interesting to see that the same procedure can behave differently on two similar networks. Another important note concerns the NMI scores during event 2. Apparently, network two was a lot harder to compute since almost every NMI score is lower than in event 1. The vastly fluctuating results are a possible indication that our study suffered from the sparsity-problem. In a follow-up study, we intend to increase the size of the networks and include ground truth by means of video information and participant recollection. Other interesting directions for future work are given by including spatial/localization information for analyzing spatio-temporal patterns [2, 33] as well as collecting more descriptive information in order to enable community detection on attributed networks, e. g., [5].

References

1. Atzmueller, M.: Data Mining on Social Interaction Networks. Journal of Data Mining and Digital Humanities **1** (June 2014)
2. Atzmueller, M., Becker, M., Kibanov, M., Scholz, C., Doerfel, S., Hotho, A., Macek, B.E., Mitzlaff, F., Mueller, J., Stumme, G.: Ubicon and Its Applications for Ubiquitous Social Computing. New Review of Hypermedia and Multimedia **20**(1), 53–77 (2014)
3. Atzmueller, M., Benz, D., Doerfel, S., Hotho, A., Jäschke, R., Macek, B.E., Mitzlaff, F., Scholz, C., Stumme, G.: Enhancing Social Interactions at Conferences. it - Information Technology **53**(3), 101–107 (2011)
4. Atzmueller, M., Doerfel, S., Hotho, A., Mitzlaff, F., Stumme, G.: Face-to-Face Contacts at a Conference: Dynamics of Communities and Roles. In: Modeling and mining ubiquitous social media, pp. 21–39. Springer (2012)
5. Atzmueller, M., Doerfel, S., Mitzlaff, F.: Description-Oriented Community Detection using Exhaustive Subgroup Discovery. Information Sciences **329**, 965–984 (2016)
6. Barrat, A., Cattuto, C., Colizza, V., Pinton, J.F., Broeck, W.V.d., Vespignani, A.: High Resolution Dynamical Mapping of Social Interactions with Active RFID. arXiv preprint arXiv:0811.4170 (2008)
7. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast Unfolding of Communities in Large Networks. Journal of statistical mechanics: theory and experiment **2008**(10), P10008 (2008)
8. Bonacich, P.: Technique for Analyzing Overlapping Memberships. Sociological methodology **4**, 176–185 (1972)
9. Brust, M.R., Rothkugel, S.: Small worlds: Strong Clustering in Wireless Networks. arXiv preprint arXiv:0706.1063 (2007)
10. Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J.F., Vespignani, A.: Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. PloS one **5**(7), e11596 (2010)
11. Coscia, M., Giannotti, F., Pedreschi, D.: A Classification for Community Discovery Methods in Complex Networks. Statistical Analysis and Data Mining: The ASA Data Science Journal **4**(5), 512–546 (2011)
12. Csardi, G., Nepusz, T.: The igraph Software Package for Complex Network Research. InterJournal **Complex Systems**, 1695 (2006), `http://igraph.org`
13. Danon, L., Diaz-Guilera, A., Duch, J., Arenas, A.: Comparing Community Structure Identification. Journal of Statistical Mechanics: Theory and Experiment **2005**(09), P09008 (2005)
14. Eagle, N., Pentland, A.S., Lazer, D.: Inferring Friendship Network Structure by Using Mobile Phone Data. PNAS **106**(36), 15274–15278 (2009)
15. Fortunato, S.: Community Detection in Graphs. Physics Reports **486**(3-5), 75 – 174 (2010)
16. Freeman, L.C.: Centrality in Social networks Conceptual Clarification. Social networks **1**(3), 215–239 (1978)
17. Girvan, M., Newman, M.E.: Community Structure in Social and Biological Networks. Proceedings of the national academy of sciences **99**(12), 7821–7826 (2002)
18. Isella, L., Romano, M., Barrat, A., Cattuto, C., Colizza, V., Van den Broeck, W., Gesualdo, F., Pandolfi, E., Ravà, L., Rizzo, C., et al.: Close Encounters in a Pediatric Ward: Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors. PloS one **6**(2), e17144 (2011)

19. Kibanov, M., Atzmueller, M., Scholz, C., Stumme, G.: On the Evolution of Contacts and Communities in Networks of Face-to-Face Proximity. In: Proc. IEEE CPSCom. pp. 993–1000. IEEE (2013)
20. Kibanov, M., Atzmueller, M., Scholz, C., Stumme, G.: Temporal Evolution of Contacts and Communities in Networks of Face-to-Face Human Interactions. Science China Information Sciences **57**(3), 1–17 (2014)
21. Macek, B.E., Scholz, C., Atzmueller, M., Stumme, G.: Anatomy of a Conference. In: Proc. ACM Hypertext. pp. 245–254. ACM (2012)
22. Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G.S., Wallinga, J., et al.: Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. PLoS medicine **5**(3) (2008)
23. Newman, M.E.: Finding Community Structure in Networks Using the Eigenvectors of Matrices. Physical review E **74**(3), 036104 (2006)
24. Newman, M.E., Girvan, M.: Finding and Evaluating Community Structure in Networks. Physical review E **69**(2), 026113 (2004)
25. Opsahl, T., Agneessens, F., Skvoretz, J.: Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths. Social networks **32**(3), 245–251 (2010)
26. Orman, G.K., Labatut, V., Cherifi, H.: On Accuracy of Community Structure Discovery Algorithms. arXiv preprint arXiv:1112.4134 (2011)
27. Raghavan, U.N., Albert, R., Kumara, S.: Near Linear Time Algorithm to Detect Community Structures in Large-Scale Networks. Physical review E **76**(3) (2007)
28. Reichardt, J., Bornholdt, S.: Statistical Mechanics of Community Detection. Physical Review E **74**(1), 016110 (2006)
29. Rosvall, M., Bergstrom, C.T.: Maps of Random Walks on Complex Networks Reveal Community Structure. PNAS **105**(4), 1118–1123 (2008)
30. Scholz, C., Atzmueller, M., Barrat, A., Cattuto, C., Stumme, G.: New Insights and Methods For Predicting Face-To-Face Contacts. In: Kiciman, E., Ellison, N.B., Hogan, B., Resnick, P., Soboroff, I. (eds.) Proc. 7th Intl. AAAI Conference on Weblogs and Social Media. AAAI Press, Palo Alto, CA, USA (2013)
31. Scholz, C., Atzmueller, M., Kibanov, M., Stumme, G.: Predictability of Evolving Contacts and Triadic Closure in Human Face-to-Face Proximity Networks. Social Network Analysis and Mining **4**(1), 217 (Aug 2014)
32. Scholz, C., Atzmueller, M., Stumme, G.: On the Predictability of Human Contacts: Influence Factors and the Strength of Stronger Ties. In: Proc. IEEE SocialCom. pp. 312–321. IEEE (2012)
33. Scholz, C., Doerfel, S., Atzmueller, M., Hotho, A., Stumme, G.: Resource-Aware On-Line RFID Localization Using Proximity Data. In: Proc. ECML PKDD. pp. 129–144. Springer, Heidelberg, Germany (2011)
34. Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.F., Quaggiotto, M., Van den Broeck, W., Régis, C., Lina, B., et al.: High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. PloS one **6**(8), e23176 (2011)
35. Wang, M., Wang, C., Yu, J.X., Zhang, J.: Community Detection in Social Networks: An In-Depth Benchmarking Study with a Procedure-Oriented Framework. Proceedings of the VLDB Endowment **8**(10), 998–1009 (2015)
36. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications, vol. 8. Cambridge university press (1994)
37. Xie, J., Kelley, S., Szymanski, B.K.: Overlapping Community Detection in Networks: The State-of-the-Art and Comparative Study. ACM CSUR **45**(4), 43 (2013)
38. Zhang, Y., Wang, L., Zhang, Y.Q., Li, X.: Towards a Temporal Network Analysis of Interactive WiFi Users. EPL (Europhysics Letters) **98**(6), 68002 (2012)