Reality Mining Via Process Mining

M. M. MohieEl-Din, M. S. Farag, O. M. Hassan

Department of Mathematics, Facility of Science,
Al-Azhar University, Cairo, Egypt
{mmeldin, farag.sayed, ohassan}@azhar.edu.eg

Abstract. Reality mining project work on Ubiquitous Mobile Systems (UMSs) that allow for automated capturing of events. Reality Mining demonstrates the power of collecting not only communication information but also location and proximity data from mobile phones over an extended period. On the other hand Process mining aims at extracting information from event logs to capture the process as it is being executed. Process mining also supports analysis of the performance of processes including Qualitative and Quantitative analysis for captured process model. This paper introduces process mining to modeling and analyzing reality mining dataset.

Keywords: Process mining; workflow mining; reality mining; ubiquitous computing; Complex social systems; User modeling; social network analysis.

1 Introduction

The Reality Mining project introduced by MIT (Massachusetts Institute of Technology) Media Laboratory to study followed 94 subjects using mobile phones preinstalled with several pieces of software that recorded and sent the researcher data about call logs, Bluetooth devices in proximity of approximately five meters, cell tower IDs, application usage, and phone status[1-2]. The collected information by 94 human subjects over the course of the academic year represent approximately 450,000 hour of information about users' location, communication and device usage behavior [3].Reality Mining demonstrate the power of collecting not only communication information but also location and proximity data from mobile phones over an extended period, and compare the resulting behavioral social network to self-reported relationships from the same group [1].

The goal of process mining is to reverse the process and collect data at runtime to support workflow design and analysis [4]. Process mining aims at extracting information from event logs to capture the business process as it is being executed [5]. The main benefit of process mining techniques is that information is objectively compiled. In other words, process mining techniques are helpful because they gather information about what is actually happening according to an event log of organization not what people think that is happening in this organization [5].

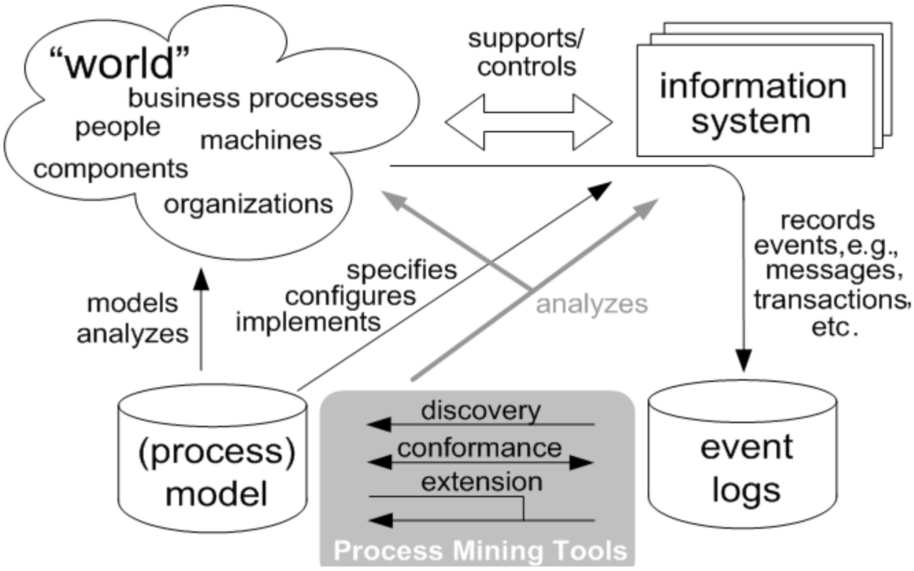


Fig. 1. Three types of process mining: (1) Discovery, (2) Conformance, and (3) Extension

In this paper we present a case study where we apply process mining techniques to modeling and analysis reality mining dataset that work on Ubiquitous Mobile Systems (UMSs). As reality mining dataset gathered from phones users' activities "event logs as called in process mining" so there are several perspectives to study this data and also there are several techniques in process mining to modeling and analysis this event logs. We will use ProM framework as a mature tools that was developed to support the various forms of process mining [6-7].

This paper structure as follows: Section 2 provides an overview about process mining concepts and techniques. Section 3 provides overview about reality mining goals and dataset structure. In section 4 we apply process mining on reality mining dataset. Section 5 concludes this paper.

2 Process Mining

Instead of starting with a process design, process mining starts by gathering information about the processes as they take place. For any information system using transactional systems or PAIS (Process Aware Information System) such as ERP (Enterprise resource planning), CRM (Customer relationship management), B2B (Business-to-business) and WFM (workflow management) systems will offer information about the order in which the events of a case are executed [5]. This information called "Event Log" and this the start point of process mining. Process mining uses the information available in this event log to reconstruct the order of activities in the form of a graphical model (i.e. process model). The model represents the executed processes based on the logs.

There are three classes of process mining techniques based on whether there is an a priori model or not as shows in Figure 1 [8]:

- Discovery: There is no a priori model and based on an event log we constructed the model.
- Conformance: There is an a priori model. This model is compared with the event log.
- Extension: There is an a priori model. This model is extended with a new aspect or perspective.

In process mining there are several techniques to discover process model. Each technique has different perspective and working strategy. Some algorithms work with local strategy to build model step by step and others work with global strategy to work based on a one strike search for the optimal model. And also there are there is differentiation between ability to extracting models dealing with noisy information, looping, duplicate tasks, incompleteness log, and also how much of information this model will be aware with. The following three different examples of process mining techniques:

- Alpha Mining [9]: this algorithm works based on local strategy technique to build model. The alpha algorithm assumes event logs to be complete and does not contain any noise. Therefore, the alpha algorithm is sensitive to noise and incompleteness of event logs. On the other hand it gives us a quick view of natural of workflow model we work on.
- Genetic Mining [10]: algorithm works based on global strategy technique to build model. This technique can deal with noisy and duplicate tasks and can provide us with detailed model. It based on genetic algorithm so we can say the time against details.
- Heuristics mining [11]: this technique extend alpha algorithm by consider the frequency of traces in the log. Heuristics miner can deal with noise, and can be used to express the main behavior.

UMSs allow for automated capturing of events. These can be used to automatically record human behavior and business processes in detail. This automated capturing of events provides an interesting application domain for process mining [12].

3 Reality Mining

Reality mining project introduced for sensing complex social systems with data collected from 94 mobile phones. Reality Mining considers Mobile phones as wearable sensors allow studying both individuals and organizations. We can divide reality mining dataset [1] into six categories as in Figure 2:

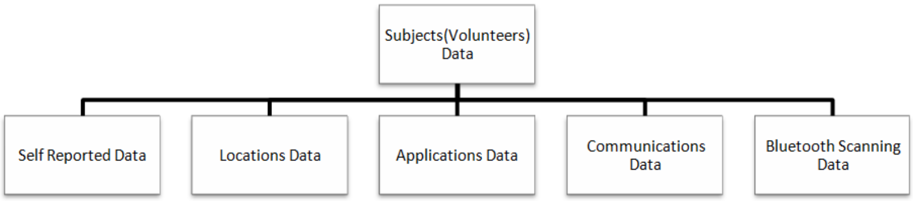


Fig. 2. Categories of data in reality mining dataset

- Subjects (Volunteers) data: related to individual personal information like working time, which group he belongs to.
- Self Reported Data: related to surveys results which represent what people think.
- Locations Data: in reality mining dataset locations associated with mobile phones cell tower IDs, as each cell tower represent unique location. So locations data related to which cell tower user belongs to during the time
- Applications Data: related to which application used during the time for each subject.
- Communications Data: related to user communications log data including type of communication (i.e. voice call, short message...), direction (i.e. Outgoing / Incoming) and duration.
- Bluetooth Scanning Data: related to observed devices by subject's mobile phone each Bluetooth scanning time.

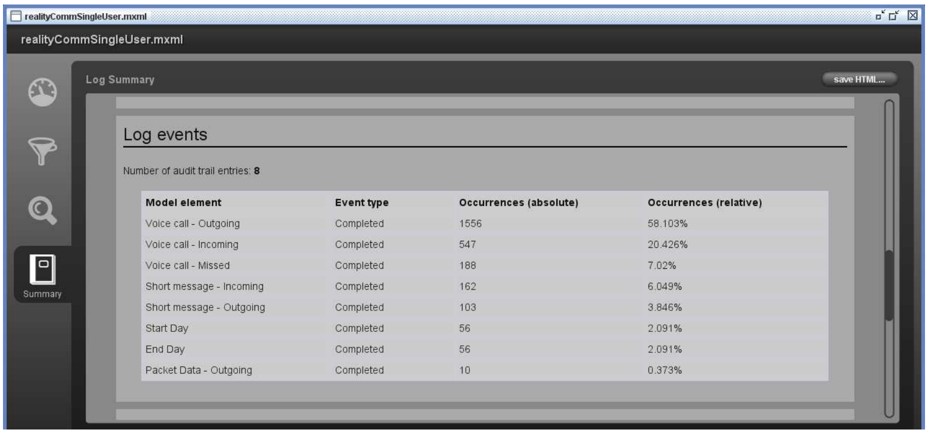
Reality mining raise interested questions related to user modeling. From reality mining dataset structure we can see that each category represent a different perspective. Self reported data represent what users think about friendship and spending time in work and so on. Others phone data can represent the actual events and relations by extracting the model for each concerned pattern.

4 Discovering Reality via Process Mining (Mining Based on Goals)

In this section we present three different patterns extracted from reality mining dataset via process mining techniques using ProM framework.

4.1 Communications Pattern

This pattern related to behavior of subjects (volunteers) to make communications using mobile phones. In reality mining dataset the communications events data form "for each volunteer" as following: (TIME) 20060720T211505 (DESCRIPTION) Voice call (DIRECTION) Outgoing (DURATION) 23 (NUMBER) 6175559821



Log events

Number of audit trail entries: **8**

Model element	Event type	Occurrences (absolute)	Occurrences (relative)
Voice call - Outgoing	Completed	1556	58.103%
Voice call - Incoming	Completed	547	20.426%
Voice call - Missed	Completed	188	7.02%
Short message - Incoming	Completed	162	6.049%
Short message - Outgoing	Completed	103	3.846%
Start Day	Completed	56	2.091%
End Day	Completed	56	2.091%
Packet Data - Outgoing	Completed	10	0.373%

Fig. 3. Summary of communications event log

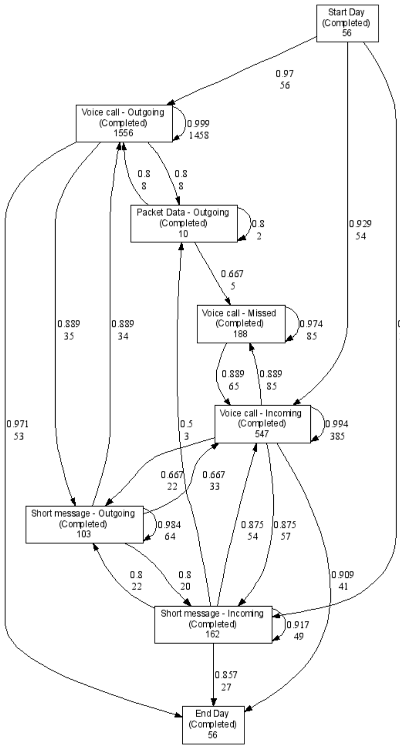


Fig. 4. Communication workflow from heuristic miner

We present one day events as a single process instance, this process instance starts with “Start Day” audit trial and end with “End Day” audit trail. The others available audit trials for each instance consisting of combination between the description and direction. By using Log Summary plug-in in ProM, the list of available audit trials in selected population as in Figure 3.

To extract the workflow model that represents the communication pattern in our selected population, we use heuristic mining [11] algorithm .Figure 4 shows the extracted model from communications log using Heuristics Miner plug-in in ProM.

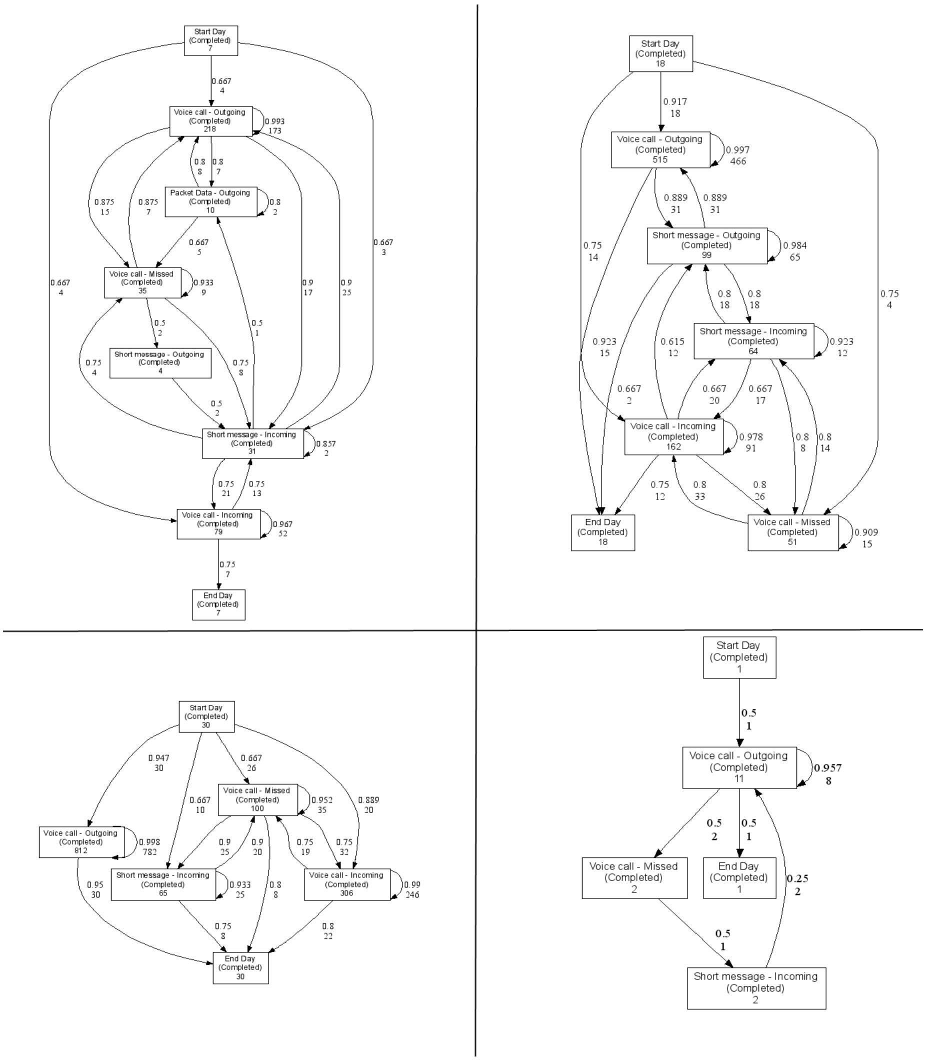


Fig. 5. Clustered models using DWS plug-in

The previous workflow represents the generic communication behavior model, but there is no commitment to hold one pattern during the time, so we use process clustering to extract the workflow schemas from the event log. We use the DWS (Disjunctive Workflow Schema) technique to clustering process. Figure 5 shows the extracted schemas using DWS Mining plug-in in ProM.

Log events

Number of audit trail entries: **26**

Model element	Event type	Occurrences (absolute)	Occurrences (relative)
Phone	Completed	11127	43.151%
ScreenSaver	Completed	4490	17.413%
Logs	Completed	3360	13.03%
Phonebook	Completed	1963	7.613%
Menu	Completed	1420	5.507%
ClockApp	Completed	964	3.738%
profileapp	Completed	713	2.765%
Pinboard	Completed	303	1.175%
context_log	Completed	266	1.032%
Start Day	Completed	264	1.024%
End Day	Completed	264	1.024%
mce	Completed	218	0.845%
Camera	Completed	212	0.822%
MediaGallery	Completed	103	0.399%
Calendar	Completed	35	0.136%
gs	Completed	18	0.07%
Browser	Completed	12	0.047%
MediaPlayer	Completed	11	0.043%
BtUi	Completed	10	0.039%
Appmngr	Completed	9	0.035%
SysAp	Completed	7	0.027%
PSLN	Completed	6	0.023%
mmcapp	Completed	4	0.016%
cshelp	Completed	3	0.012%
Videorecorder	Completed	3	0.012%
Appinst	Completed	1	0.004%

Fig. 6. Summary of mobile applications usage

4.2 Mobile usage pattern

This pattern related to usage of application on mobile during the time. Applications event logs record which application opened in which time. In this pattern also we consider one day events as a single process instance. Each process instance starts with "Start Day" audit trial and end with "End Day" audit trial. First quick statistical analysis of applications log by extraction what applications used and the Occurrences percentage for each one. We use Log Summary plug-in to extract Figure 6 for our selected population.

In mobile usage pattern we focus on relativity of applications usage and distribution of applications usage during the time. To extract model represent relativity of applications usage we use fuzzy miner [13]. Fuzzy miner focuses on clustering model by aggregate the high relative events. Also the Fuzzy Miner offers a dynamic view of the process by replaying the log in the model. Figure 7 shows the extracted model using Fuzzy Miner plug-in that is representing the applications usage pattern. To extract the distribution of application usage we use the Dotted Chart analysis plug-in as shows in Figure 8. From dotted chart it is easy to know in which time user focus on each application.

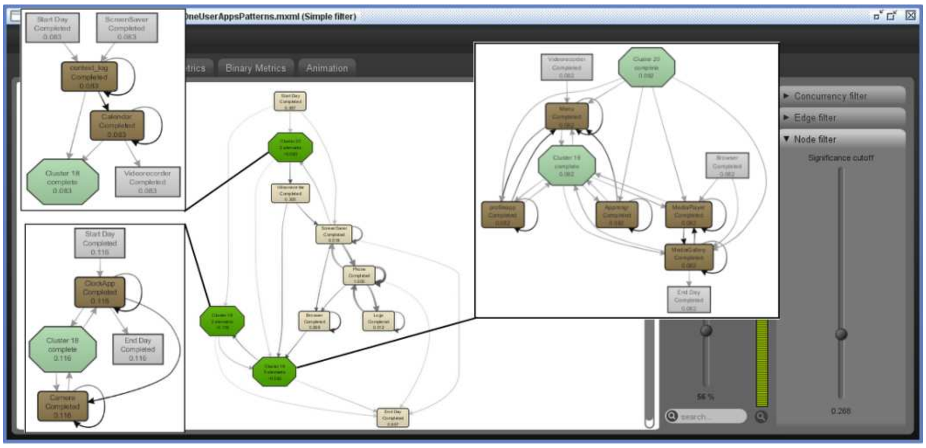


Fig. 7. Mobile applications usage model using fuzzy miner

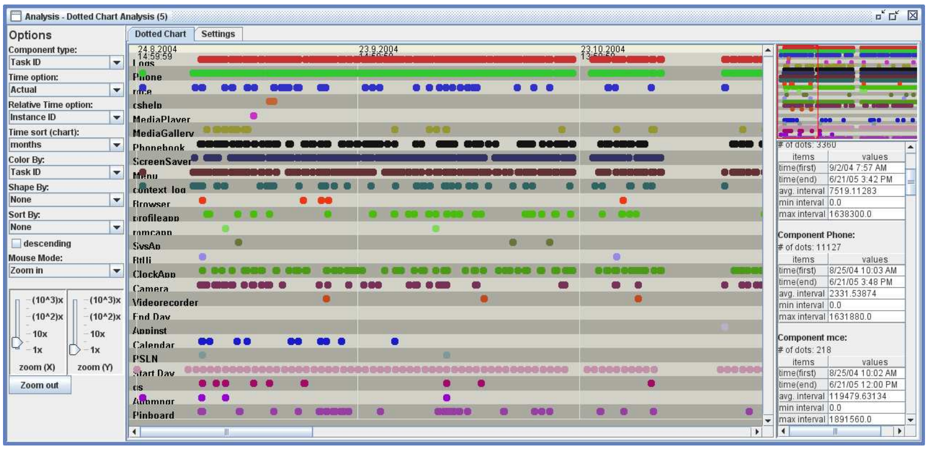


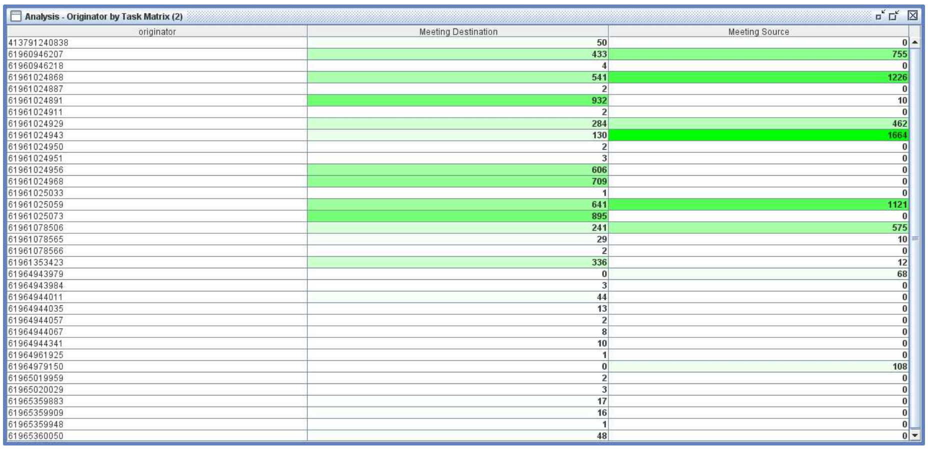
Fig. 8. Dotted chart for mobile applications usage model

4.3 Relationship pattern

One of most interested raised issue from reality mining is “What Does Friendship Look Like?” and what different between self reported friendship and actual friendship [1]. In reality mining project each mobile phone scan the environment for Bluetooth devices once every 5 min, and the Bluetooth log data form as following: (TIME) 20060721T111222 (devices): 000e6d2a3564 [Amy’s Phone], 000e6d2b06ea [Jon’s Palm Pilot]

In this section we use Bluetooth logs as social activities to extract the subjects’ relationship pattern, so our main focus not the activities itself but the originators of these activities. We present each detected device by subjects’ mobile as a single process instance. Each process instance consist of two audit trials, first one represent

"Meeting Source" with subject himself as originator for this audit trial, and the second audit trial represent "Meeting Destination" with detected device as originator for this audit trial.



Analysis - Originator by Task Matrix (2)

originator	Meeting Destination	Meeting Source
413791240838	50	0
61960946207	433	755
61960946218	4	0
61961024868	541	1226
61961024887	2	0
61961024891	932	10
61961024911	2	0
61961024929	284	462
61961024943	130	1664
61961024950	2	0
61961024951	3	0
61961024956	606	0
61961024968	709	0
61961025033	1	0
61961025059	641	1121
61961025073	895	0
61961078506	241	575
61961078565	29	10
61961078566	2	0
61961353423	336	12
61964943979	0	68
61964943984	3	0
61964944011	44	0
61964944035	13	0
61964944057	2	0
61964944067	8	0
61964944341	10	0
61964961925	1	0
61964979150	0	108
61965019959	2	0
61965020029	3	0
61965359883	17	0
61965359909	16	0
61965359948	1	0
61965360050	48	0

Fig. 9. Bluetooth (scanner/detected) devices frequency

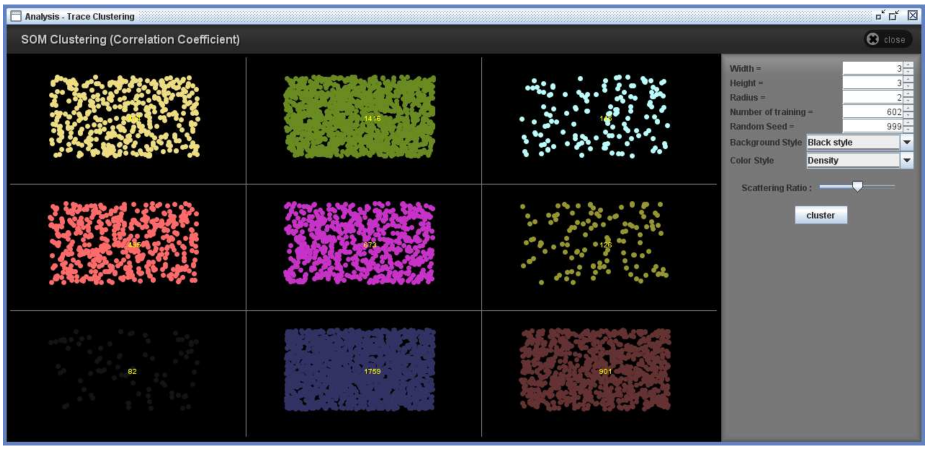


Fig. 10. Clustering of Bluetooth scanning process using Trace Clustering plug-in

We use the Originator by Task Matrix plug-in to list the frequency of showing each subject as Bluetooth scanner or as detected device as shows in Figure 9. Also this data conditionally highlighted based on frequency number.

Before extracting relationship model we check how our selected population data connected with each other. By using clustering analysis we can grouping related data and take an overview about our population. We use Trace Clustering plug-in to clustering events, Trace Clustering plug-in support multi clustering algorithm, we use SOM algorithm and Correlation Coefficient distance function to grouping data, as shows in Figure 10, This represent scanning process clustering itself not subjects relationship clustering.

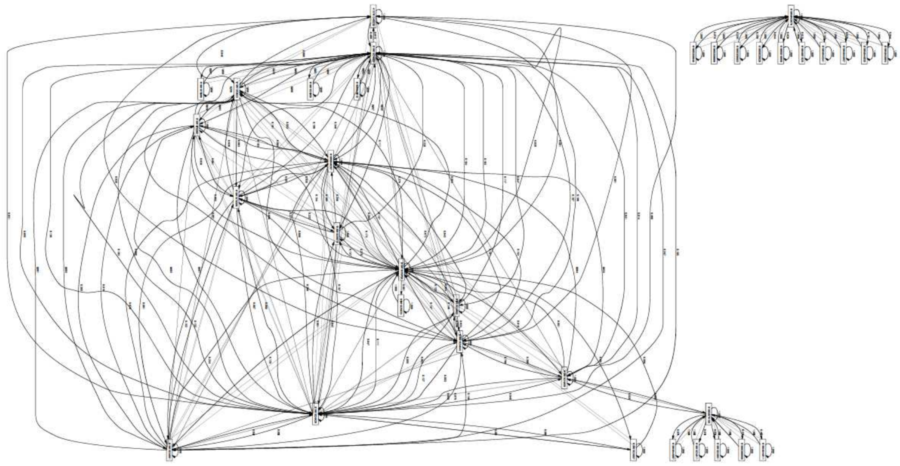


Fig. 11. Extracted social network model from Bluetooth scanning process

There are several techniques in process mining that address the social activities mining, e.g., organizational mining, social network mining, mining staff assignment rules, etc. [14]. In this pattern we use Social Network Miner plug-in to model the originators relationship. There are several techniques to analysis social networks. Figure 11 shows the model obtained by applying Working Together method to analysis the social network. It is clear enough to show that there are high related people "i.e. showing in groups" and also there are several subjects represent a central points.

5 Conclusions

This paper focuses on applicability of process mining in the UMS systems. For our case study, we have used data coming from reality mining dataset. We apply process mining to analyzing and modeling the patterns of reality mining dataset.

Reality mining dataset is reachable and accurate enough to have different realistic patterns with different perspectives. And the Varity on process mining modeling and analysis techniques Allow to analysis event logs with different perspectives based on goal of analysis process.

Although that process mining work with event logs with ordered audit trails, there is no restriction to use it to extract patterns with no dependency on events order like social activities. But With large size population we face limitation in clustering process techniques in process mining, despite the smallest population included in largest population but process clustering provides mature result just with small population.

6 References

1. Eagle, N., Pentland, A., Lazer, D.: Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences 106, 15274-15278 (2009)
2. http://reality.media.mit.edu
3. Eagle, N., Pentland, A.: Reality mining: sensing complex social systems. Personal and Ubiquitous Computing 10, 255-268 (2006)
4. Van der Aalst, W., Van Dongen, B., Herbst, J., Maruster, L., Schimm, G., Weijters, A.: Workflow mining: A survey of issues and approaches. Data & Knowledge Engineering 47, 237-267 (2003)
5. Van der Aalst, W., Weijters, A.: Process mining: a research agenda. Computers in Industry 53, 231-244 (2004)
6. van der Aalst, W., van Dongen, B., Günther, C., Mans, R., de Medeiros, A., Rozinat, A., Rubin, V., Song, M., Verbeek, H., Weijters, A.: ProM 4.0: Comprehensive support for real process analysis. Petri Nets and Other Models of Concurrency–ICATPN 2007 484-494 (2007)
7. http://prom.win.tue.nl/research/wiki
8. Mans, R., Schonenberg, M., Song, M., Aalst, W., Bakker, P.: Application of Process Mining in Healthcare–A Case Study in a Dutch Hospital. Biomedical Engineering Systems and Technologies 425-438 (2008)
9. Van der Aalst, W., Weijters, T., Maruster, L.: Workflow mining: Discovering process models from event logs. IEEE Transactions on Knowledge and Data Engineering 1128-1142 (2004)
10. de Medeiros, A., Weijters, A., van der Aalst, W.: Genetic process mining: an experimental evaluation. Data Mining and Knowledge Discovery 14, 245-304 (2007)
11. Weijters, A., van der Aalst, W., de Medeiros, A.: Process mining with the heuristics miner-algorithm. Technische Universiteit Eindhoven, Tech. Rep. WP 166, (2006)
12. De Medeiros, A., Van Dongen, B., Van Der Aalst, W., Weijters, A.: Process mining for ubiquitous mobile systems: an overview and a concrete algorithm. Ubiquitous Mobile Information and Collaboration Systems 151-165 (2004)
13. Van der Aalst, W., Giinther, C.: Finding structure in unstructured processes: The case for process mining. In: Conference Finding structure in unstructured processes: The case for process mining, pp. 3-12. (Year)
14. Van Der Aalst, W., Reijers, H., Song, M.: Discovering social networks from event logs. Computer Supported Cooperative Work (CSCW) 14, 549-593 (2005)