

Social Network and Geospatial CS419/519 Information Visualization

Joel Lonbeck and David Burri

Abstract—The VAST Challenge is offered every year as part of the Infovis conference. The contest has the purpose of pushing the forefront of visual analytics tools using benchmark data sets and establishing a forum to advance visual analytics evaluation methods. The challenge consists of one or more mini Challenges which can be solved independently and can result, by bringing them together, in the Grand Challenge. We have chosen the "Social Network and Geospatial" mini challenge from 2009 in which officials have received intelligence that a terrorist group operating within a social network have recruited a government employee. Designing an advanced interactive visualization will allow the terrorist social hierarchy to be discovered within the social network, and the identity of key people of interest to be discovered.

Index Terms—InfoVis, VAST Challenge, Social Network, Network Graph.

I. INTRODUCTION

EMBASSY employees are known to have used the social networking/micro-blogging tool, Flitter, to communicate with colleagues and friends. However, the Flitter network has also been used by a criminal ring that may have recruited an employee. We have been provided with Flitter logs and have been instructed to analyze the data in order to determine the identities of key contacts within the criminal ring.

The Flitter logs consists of three tab-delimited files. The first log file describes a list of entities, which may define users, cities, or countries. The second log file contains communication links between the users as well as geographical links that match cities to countries. The last log file describes the geographical location of users by defining user-to-city links. The provided dataset includes 6000 people from 12 different cities out of 4 countries whereby there are 29'888 person-to-person connections. A map of Flovania has also been provided and includes details regarding its major cities and information about neighboring countries. The datasets used for these challenges are synthetic: that is, they

are a blend of computer- and human-generated data and should not be regarded as representative of any geographical location, people group, or social network.

Figure ?? depicts on the left side the table with the entities and on the right side on the top an example of the person-to-city and on the bottom the person-to-person link table.

ID	Name	Type	ID	City
INT	STRING	STRING	INT	STRING
1	@irvin	person	1	Koul
2	@rachel	person	2	Kouvnic
3	@dykemaperson		3	Koul
...			4	Kouvnic
6011	Transpasko	city	...	
6012	Tulamuk	city		
6013	Flovania	country	ID1	ID2
6014	Trium	country	INT	INT
...			1	2
			3	1
			3	2
			...	

Fig. 1. Excerpts from Flitter log files

A. Task 1

The client believes that an employee communicated with his/her handler(s) through Flitter network, however they do not know the Flitter name of the handler nor of the espionage organization. But they believe that the associated network may take one of two forms of social structures described as follows:

A. The employee has about 40 Flitter contacts. Three of these contacts are his "handlers", people in the criminal organization assigned to obtain his cooperation. Each of the handlers probably has between 30 to 40 Flitter contacts and share a common middle man in the organization, who is code-named Boris. Boris maintains contact with the handlers, but does not allow them to communicate among

themselves using Flitter. Boris communicates with one or two others in the organization and no one else. One of these contacts is his likely boss, who is code-named Fearless Leader. Fearless Leader probably has a broad Flitter network (well over 100 links), including international contacts. See ??.

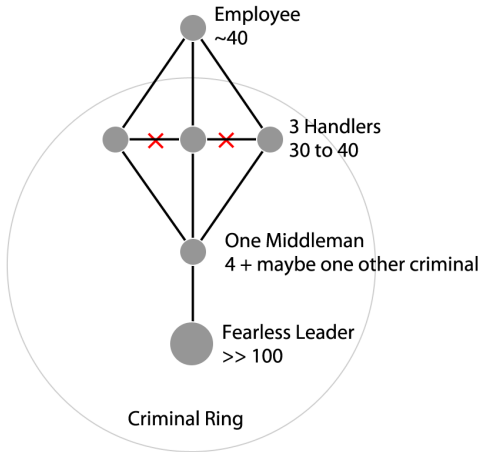


Fig. 2. Network Structure A

B. The employee has about 40 Flitter contacts. Three of these contacts are his "handlers", people in the organization assigned to obtain his cooperation. Each of the handlers likely has between 30 to 40 Flitter contacts, and each probably has his or her own middle man in the organization, who are code-named Boris, Morris and Horace. It is probable the middle men will not allow the handlers to communicate among themselves using Flitter. Each of the middle men probably communicate with one or two others in the organization, and no one else. One of the contacts for all of the middle men is the head of the organization, Fearless Leader. Fearless Leader has a broad Flitter network (well over 100 links) including international contacts. See ??.

Thus the first task is to search for the closest match scenario and provide the social network structure of the criminal ring.

II. RELATED WORK

In the article "Visualizing Social Networks", Linton Freeman reviews the long history of images used in the field of visualizing social networks. He discusses work from early hand-drawn images up to the use of interactive visualizations whereby the authors emphasis is on the use of graphs based on points and lines. The article concludes with predicting more use of animated techniques using

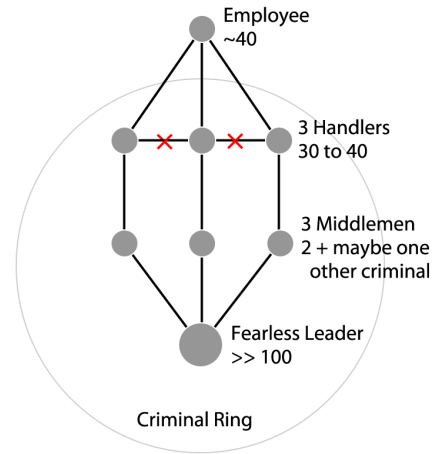


Fig. 3. Network Structure B

forms of spring embedders to place the points and more interactivity as well. [?]

Anthony Dekker discusses in the paper "Conceptual Distance in Social Network Analysis" an approach to social network analysis based on a concept of distance between people. The distance between two people can be any encoding of the relationship e.g. how often they communicate or the cultural distance and so on. In the conclusion the author claims that the introduced distance is a concept that the human brain is skilled at thinking about and visually judging. [?]

Another source of a lot of related work is called VisualComplexity.com. VisualComplexity.com is a visual exploration on mapping complex networks founded by Manuel Lima. This website is intended to be a unified resource space for anyone interested in the visualization of networks. [?]

III. DESIGN

Can we identify a criminal network structure?

Which of the two given social structures, A or B, most closely match the scenario you have identified in the data?

What role does geographic location play in the social network and in the network of the criminals?

These main questions stated by the problem statement can be divided into more specific questions such as: What is the name of the suspected employee? What are the names of the persons and the positions they hold within the criminal network? And so on. The goal of this chapter is to introduce

our design aimed to answer these questions. For a more elaborated question section see our design justification document.

A. Filtering

Since the given dataset is large and the problem statement tends to be solvable by filtering the data, we introduce a filtering technique that may be applied to the data set. Our main design allows a user to add as many filters as desired. Relationships may be defined between two filters to help facilitate the logical ordering of the social hierarchy. That is, a filter can search for a node degree range as well as the number of required connections to a previously filtered set of nodes. Defining a range of required connections between two filters will also act as an intersection of the two filtered data sets, and provide the user with a narrower set of data that matches their specifications.

B. Data Encoding

Through the design process we decided to use a standard network graph based on the following arguments: A network graph can show connections very efficiently so long as there is no need to display the whole graph – See Section ???. Filters can be implemented by adding Bubbles these are groups of nodes and the comparison is easy of the network structure to the given structure. Furthermore most users are very familiar with this type of visualization.

Table ?? shows the data given by the problem statement and the filtering as well as the encoding we chose.

Variable	Type	Encoding
User	N	Circle / Spatial Position / Color for Selection
User Name	N	Text
Connection	N	Line Between Nodes
# Connections	Q	Node Size / Text
Filter	O	Circle / Text / Group of Nodes

TABLE I
VARIABLES AND ENCODING

The nodes are placed spatially using spring embedders achieved with a physical particle system. Encoded in the size of the node is the degree that is the number of connections this node has

to other nodes. Additionally the stroke of the node is highlighted with red if the mouse is over and by double clicking the node color changes to red and the name as well as the degree will be shown. To give a user more flexibility the nodes can be dragged which is sometimes useful to unravel the graph. Arcs between the nodes have a fixed stroke with since they carry no encoding.

Since the visualization has detailed and overview information as well we chose to use a zoom and pan interface to allow a user to have a better understanding of the information presented and most of users are familiar with this type of interface.

Filters are formed to what we call Bubbles. Every filter is placed on a fixed position denoted with a circle. All nodes corresponding to this filter are grouped dynamically around this circle. These circles can be moved to provide an interface that is as flexible as possible.

IV. IMPLEMENTATION

how we built everything.

A. Database

stuff about database

V. RESULTS

the results of our work

VI. CONCLUSION

The conclusion goes here.

ACKNOWLEDGMENT

The authors would like to thank...

REFERENCES

- [1] Linton C. Freeman, *Visualizing Social Networks*, Journal of Social Structure, JoSS Article: Volume 1, 2000.
- [2] Jeffrey C. Johnson and Lothar Krempel, *Network Visualization: The "Bush Team" in Reuters News Ticker 9/11-11/15/01*, Journal of Social Structure, (Vol. 5, No. 1), 2004.
- [3] Anthony Dekker, *Conceptual Distance in Social Network Analysis*, Journal of Social Structure, (Vol. 6, No. 3), 2005.
- [4] Peter Hoek, *Parallel Arc Diagrams: Visualizing Temporal Interactions*, Journal of Social Structure, (Vol. 12, No.7), 2011.
- [5] Manuel Lima, *Visual Complexity*, www.visualcomplexity.com, 2012.
- [6] VisWeek 2009, IEEE VAST Challenge 2009, <http://hcil.cs.umd.edu/localphp/hcil/vast/index.php/taskdesc/index>, 2009.