Visualization of Uncertainty and Analysis of Geographical Data

VAST 2009 Flitter Mini Challenge Award

Jo Wood* Aidan Slingsby† Naz Khalili-Shavarini‡ Jason Dykes§ David Mountain¶

the giCentre (http://gicentre.org), School of Informatics
City University London

1 Introduction

A team of five worked on this challenge to identify a possible criminal strucutre within the Flitter social network. Initially we worked on the problem individually, deliberately not sharing any data, results or conclusions. This maximised the chances of spotting any blunders, unjustified assumptions or inferences and allowed us to triangulate any common conclusions. After an agreed period we shared our results demonstrating the visualization applications we had built and the reasoning behind our conclusions. This sharing of assumptions encouraged us to incorporate uncertainty in our visualization approaches as it became clear that there was a number of possible interpretations of the rules and assumptions governing the challenge. This summary of the work emphasises one of those applications detailing the geographic analysis and uncertainty handling of the network data.

2 Geographical Mapping

City locations were digitized and base mapping redrafted using LandSerf (`www.landserf.org`) to make geographic interpretation easier (see Figure 1). Coordinates of city locations were added to a normalised MySQL database holding tables `People`, `Cities` and `Links`. This allowed a range of geographic analysis tasks to be performed on the data.

To assess the general geographic patterns of Flitter members and their contacts, a *flow map* [1] of all Flitter contacts was produced (Figure 1) where the thickness of lines between any two cities is proportional to the numbers of Flitter users in contact with each other in those cities. This showed that the majority of contacts were within Flovania and concentrated between the three cities of Koul, Prounov and Kouvnic. It also revealed there were at least some contacts between every pair of cities in the study region.

To explore the inter-city connections in greater depth, we constructed an *Origin-Destination (OD) map* [2] showing the numbers of Flitter connections between every pair of cities arranged in approximate geographic position (see Figure 2). Each large square of the OD map contains all the Flitter connections from the city at that location (e.g. top-left square represents Otello). Each of the smaller colored squares represents a connection with each of the cities in the study region. This provided us with a base distribution with which to compare the geography of the criminal network.

Filtering out Flitter members that did not conform to the known characteristics of the criminal network allowed us to establish a relatively small number of candidate networks. The geography of these candidates helped us to establish the most likely core criminal network (see Figure 3). The target (Schaffter) and handlers (Reitenspies, Pettersson, Kuishnir) were based in the same large city (Prounov). This confirmed that "*target may be in a large city*" and assists face-to-face interaction between target and handlers. It supported the rejection of the Lafouge-(Lonning, Formenti, Krinz) and Supornpaibul-(Gusat, Letelier, Bailey) networks, having protagonists in multiple cities including a smaller one (Sresk). The central location of middleman (Good) in Kanvic is consistent with "*middleman might be in nearby smaller location*". Kannvic is a small city and close to Prounov and its central location assists the middleman's role of intermediary between protagonists. It supported the rejection of potential middlemen (Cuatto and Rosch) who are located in a larger city (Koul). The leader (Szemeredi) is less vulnerable to discovery in a smaller city (Kouvnic), but "*require[s] a presence in a larger city*" via Kroon, Krupp and Quisquater in Koul, fulfilling international contact middlemen roles as part of a fledgling 'Scenario A' network. The entire criminal network formed a 'northwestern-axis' comprising Kouvnic, Otello, Kannvic, Prounov and Koul.

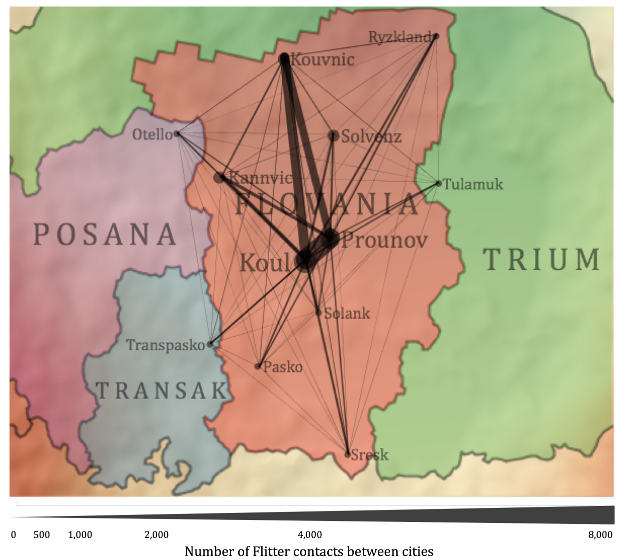

Figure 1: Flitter city connections produced using FlowMappa

3 Incorporating Uncertainty

Although the suspected criminal network was characterised by a number of rules (e.g. "*Each of the handlers probably has between 30 to 40 Flitter contacts and share a common middle man in the organization*"), the use of words such as "may", "probably" and indeed uncertainty as to whether the actual network confirmed to these suspected rules meant that rules could not be relied upon to identify the network. We therefore incorporated classification uncertainty into our visualisation allowing for fuzzy role classifica-

*e-mail: jwo@soi.city.ac.uk
†e-mail: a.slingsby@soi.city.ac.uk
‡e-mail: nazanin.khalili-shavarini.1@city.ac.uk
§e-mail: jad7@soi.city.ac.uk
¶e-mail: dmm@soi.city.ac.uk

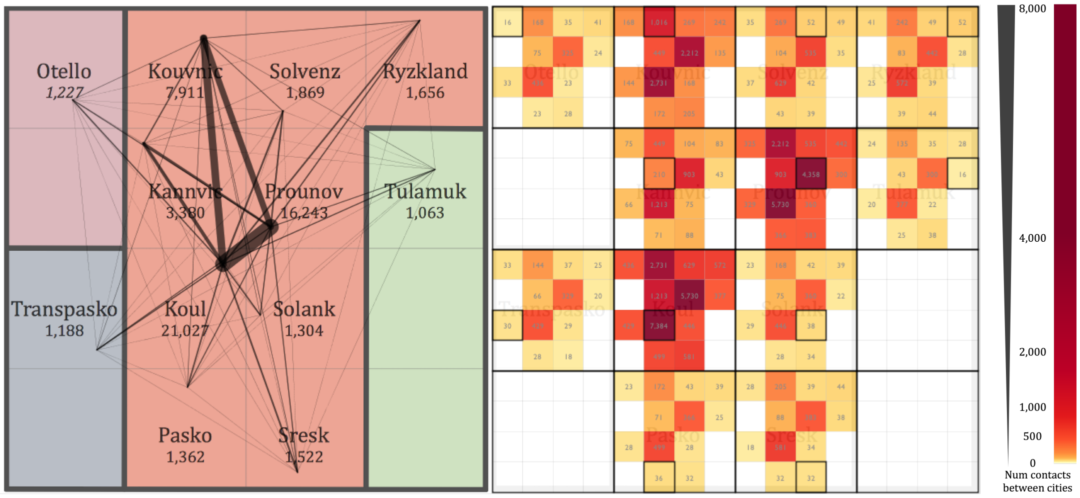

Figure 2: *Left*: Flow map of inter-city Flitter contacts. Number indicates number of Flitter members in each city. *Right*: Origin-Destination map showing numbers of contacts between all combinations of cities.

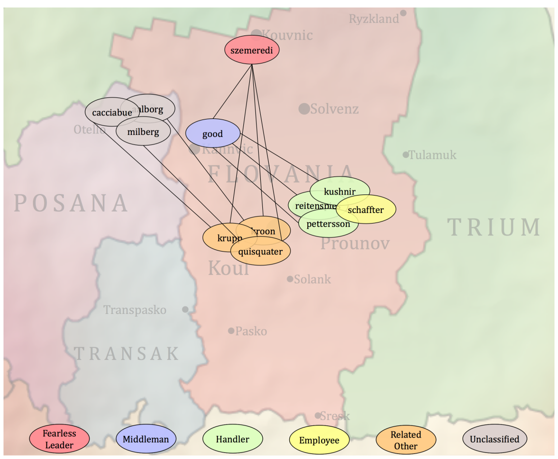

Figure 3: Geographic distribution of the criminal network

tion. Each Flitter user was given a fuzzy membership value for each role depending on their number of contacts. Users were represented using colored ellipses according to their role, using transparency to allow multiple roles and sized according to membership value (see Figure 4).

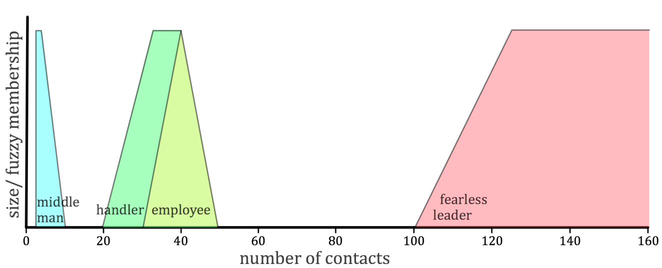

Figure 4: Node size rules for categorized Flitter users

Ellipses were positioned using a spring embedded layout where attraction was determined by the connections between nodes of different roles (see Figure 5). Similar roles were clustered, employees spaced from handlers by a unit distance and from middlemen by two unit distances etc. This allowed graph location to be used to assess network uncertainty (e.g. potential middlemen who were not connected to likely handlers were separated from those who were). The layout allowed manual repositioning and querying of ellipses in order to support hypothesis generation.

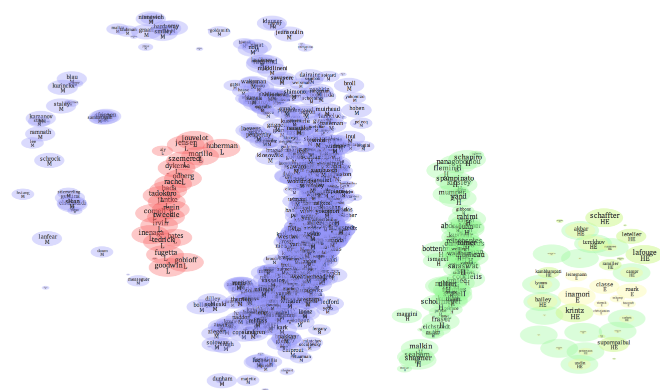

Figure 5: Classification uncertainty represented by node size. Note that some Flitter users can have multiple classes such as Employee(E) and Handler(H).

By incorporating geographic position and uncertainty into our visualization process we were able both to reject a number of possible candidate networks as well as assess the likelihood that our proposed network did indeed match the probable characteristics of a criminal network. Research group members working independently with different graphical applications and techniques before combining results and conclusions allowed us to triangulate results and assess our assumptions about the criminal network.

References

[1] W. Tobler. Experiments in migration mapping by computer. *The American Cartographer*, 14(2):155–163, 1987.

[2] J. Wood, J. Dykes, A. Slingsby, and R. Radburn. Flow trees for exploring spatial trajectories. In D. Fairbairn, editor, *Proceedings of the GIS Research UK 17th Annual Conference*, pages 229–234. University of Durham, Durham, 2009.