(temporal) Motifs in discussions (trees)

Alberto Lumbreras

February 17, 2016

Overview

Reddit dataset

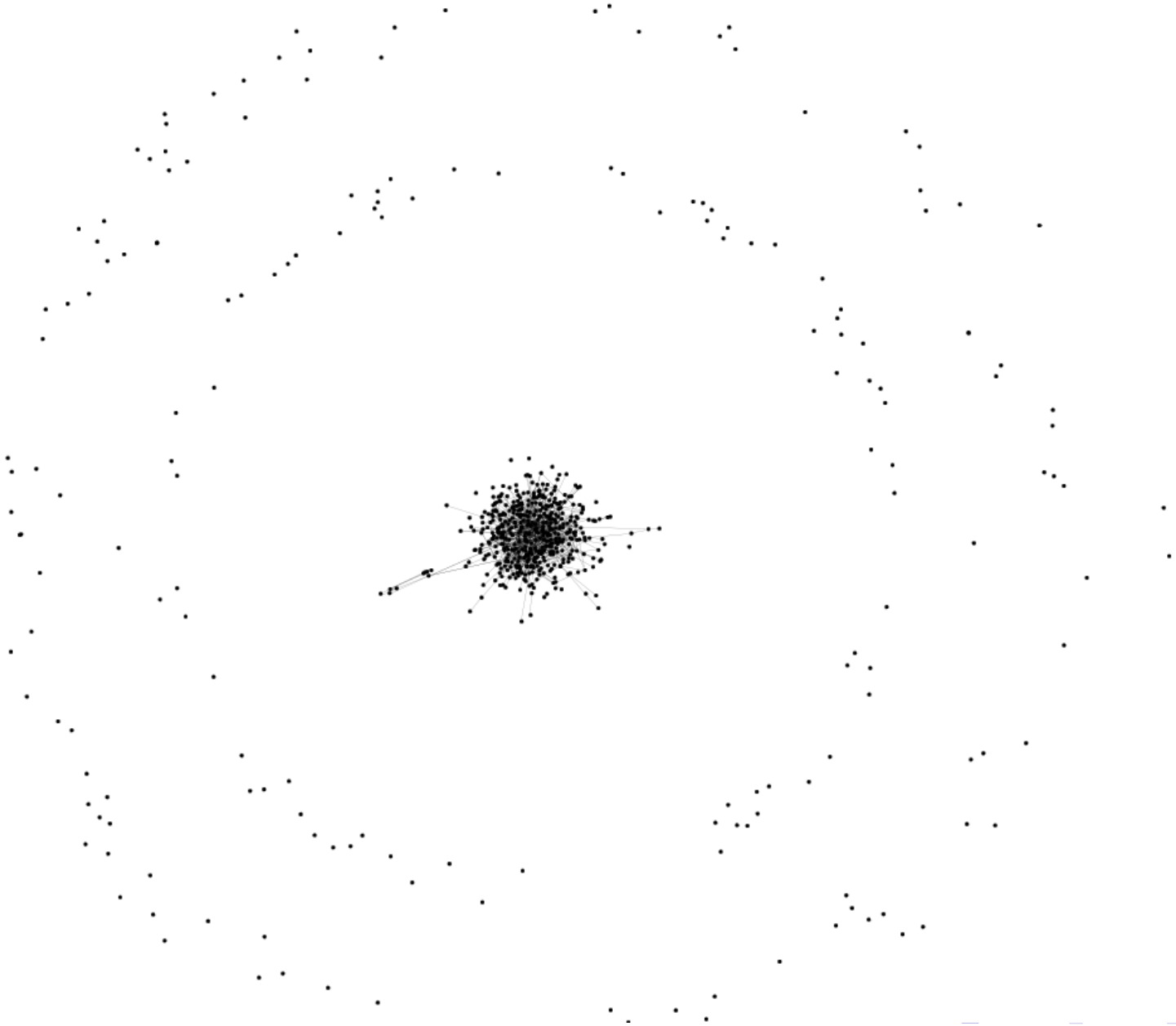
A forum of forums

Download from:
http://couch.whatbox.ca:36975/reddit/comments/monthly/

Extract forum of interest:
www.reddit.com/r/science
www.reddit.com/r/france
www.reddit.com/r/sociology
www.reddit.com/r/complexsystems
www.reddit.com/r/podemos
...

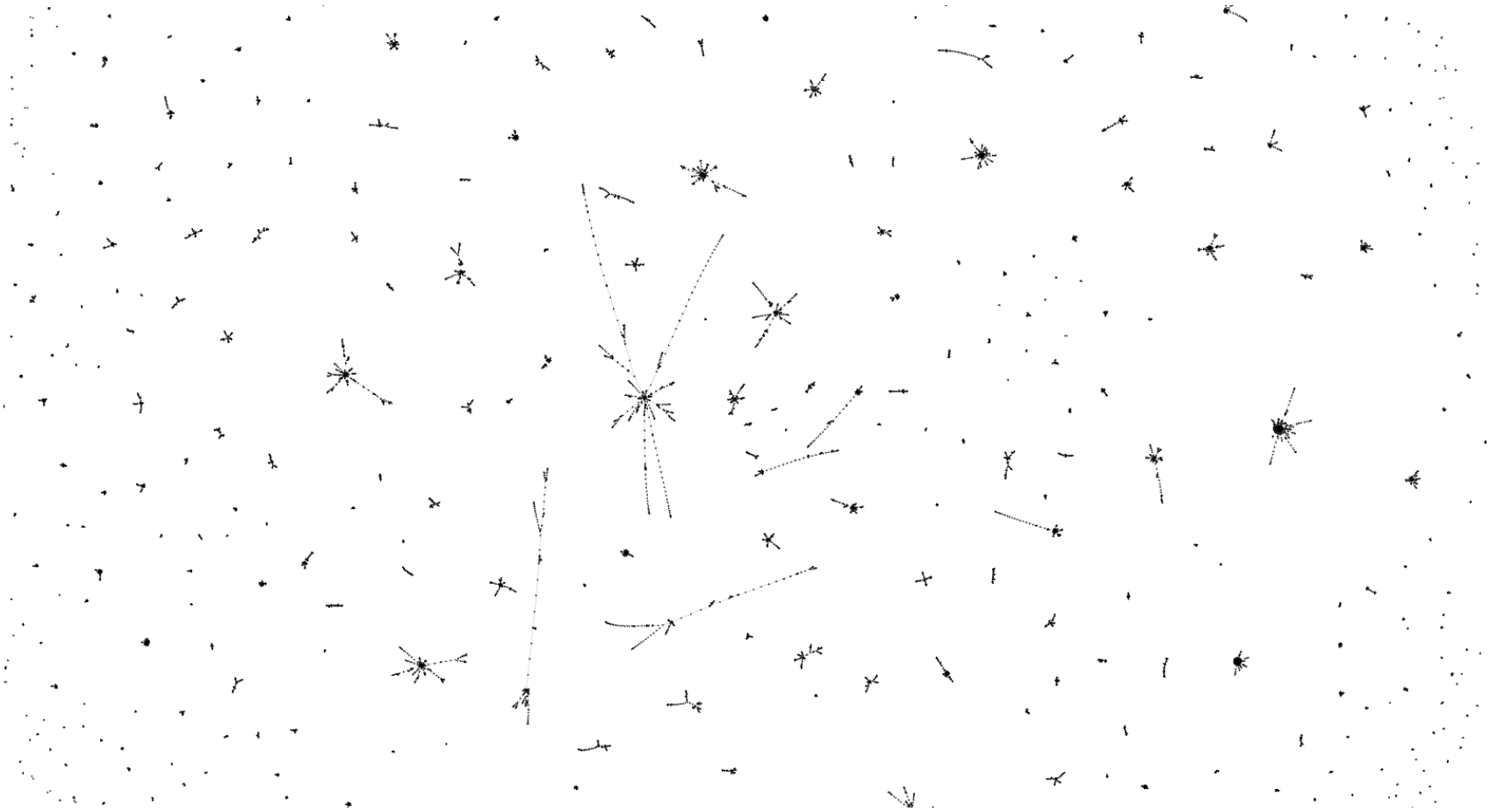
Graph representations

Graph of user interactions (a social network)



Graph representations

Trees of posts



Dynamics of conversations

Triads are not enough

Triads in **trees of posts**:

- Only 3 possible triads (dyad, chain and star)

Motif															
Motif ID			36	164	12	14	6	78	38	174	166	46	238	102	140

Triads in graph of **user interactions**:

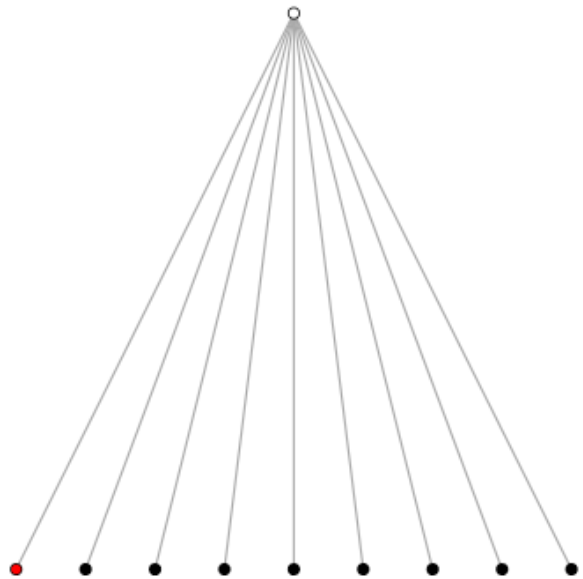
- Need of time in edges.

Temporal neighborhoods in trees

Definition

- $N_G(i, d)$: neighborhood of post i at distance d.
- $N_G(i, d, n)$: keep only the n neighbors in $N_G(i, d)$ that are *temporally* closest to post i (computed as $|t - t_i|$)
- Keep only those posts in $N_G(i, d, n)$ that have a path to i.

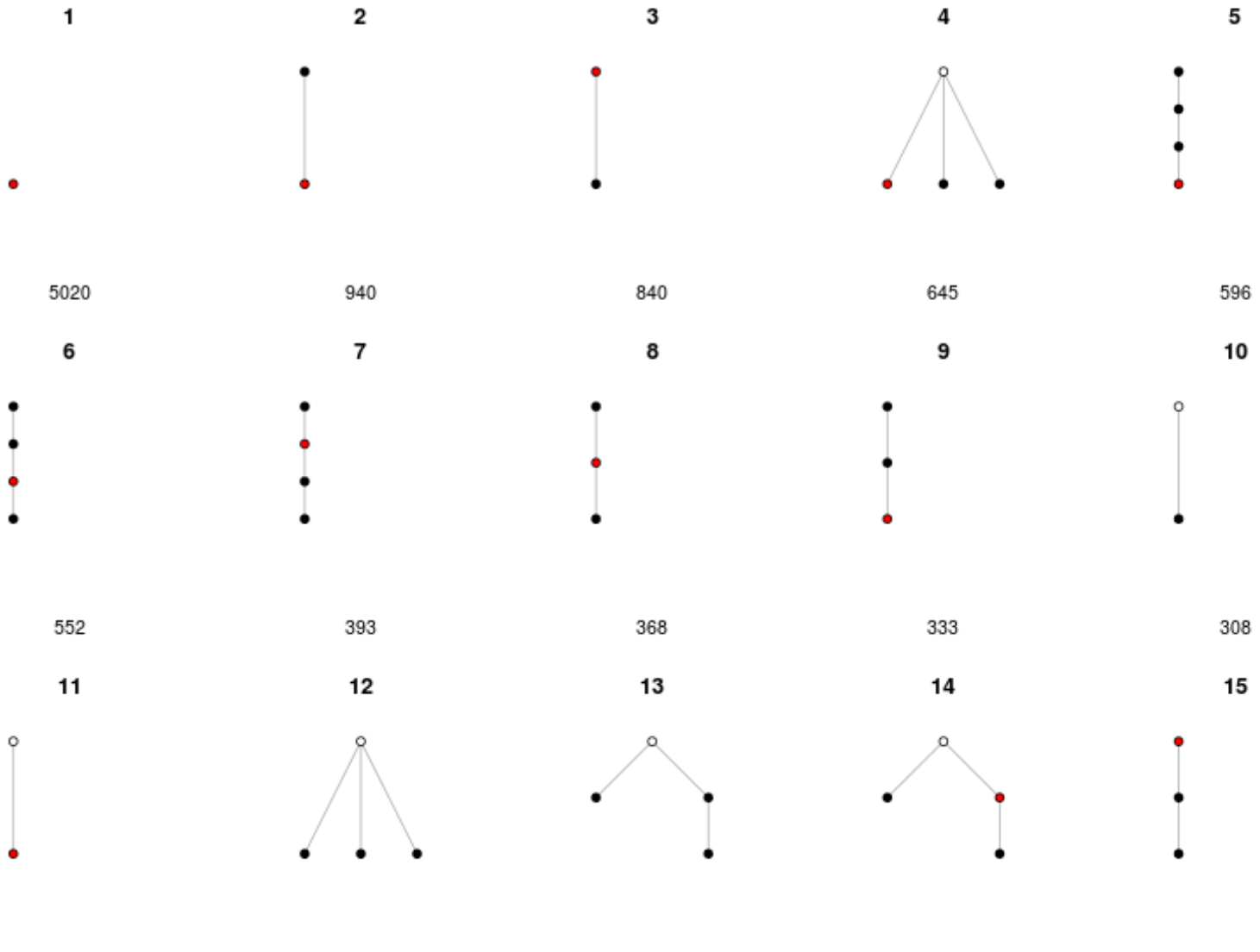
$N_G(i, d, n)$ is the **temporal neighborhood** of post i with distance d and order n. [1]



Neighborhood census

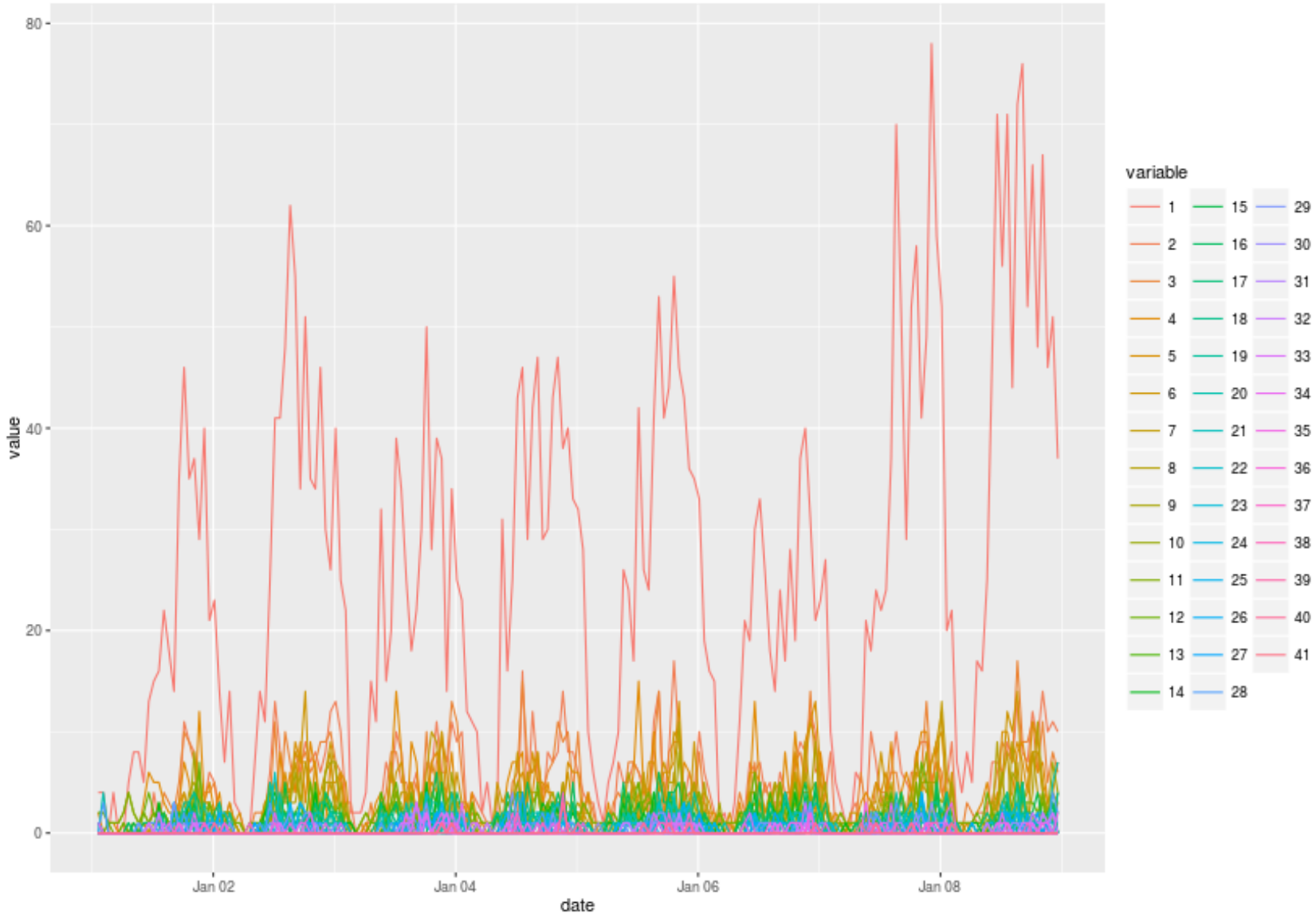
A real example

Distance $d = 4$ and order $n = 4$. 41 discussion patterns:



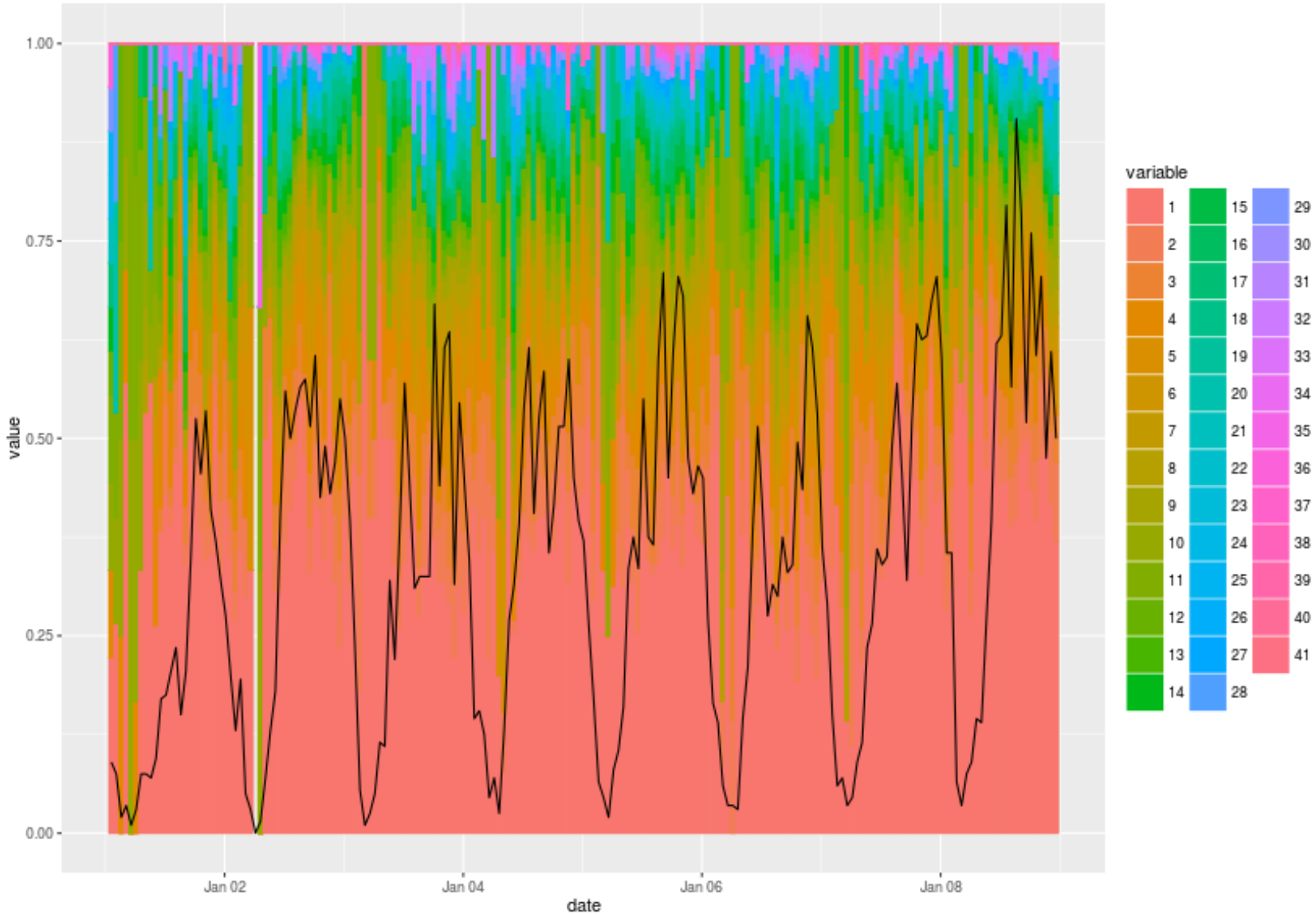
Neighborhood census

Cyclic dynamics



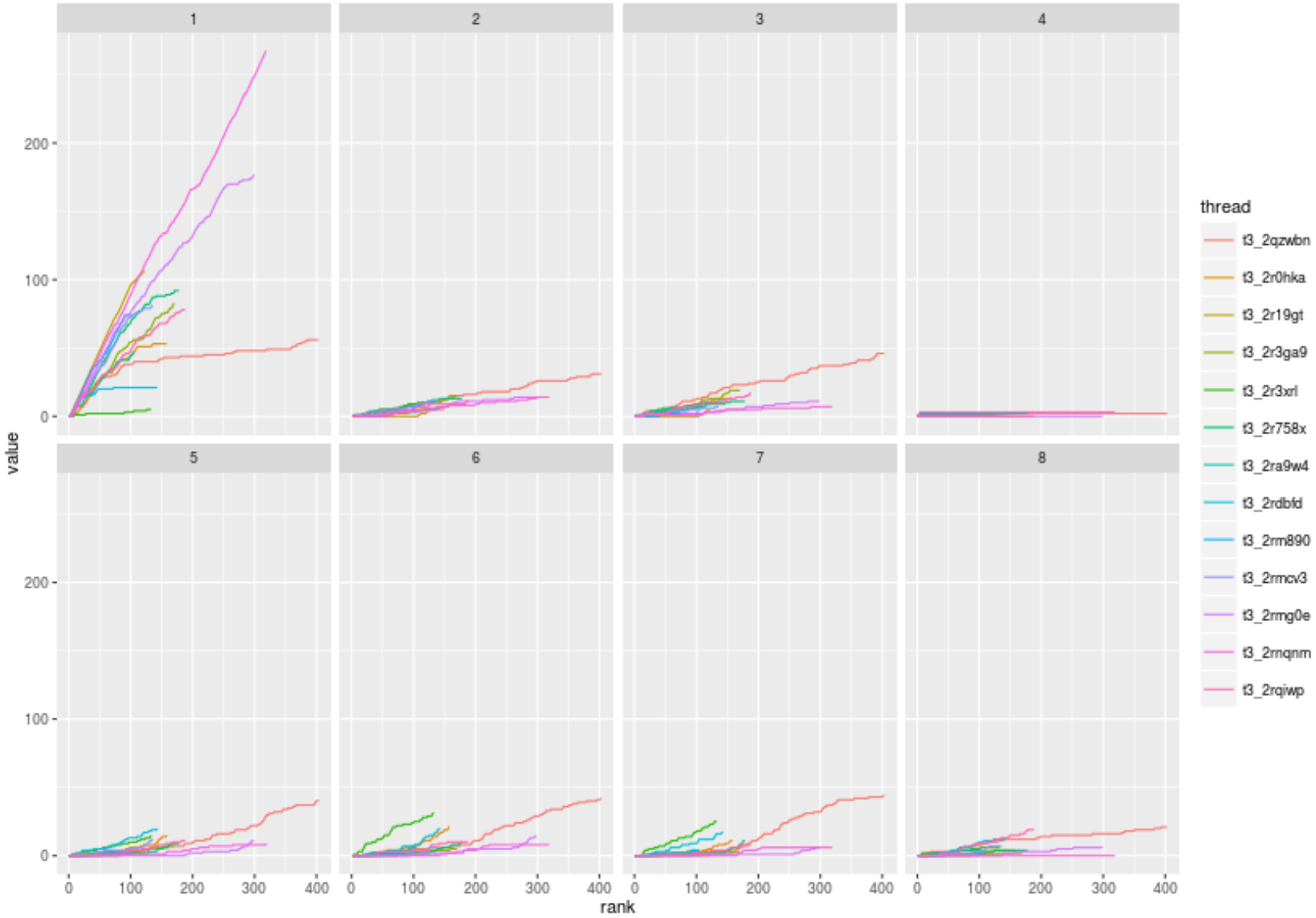
Neighborhood census

Proportions



Neighborhood census

Census and thread growth



Neighborhood census

Predictions

Q: Can we predict whether a thread will succeed based on its initial structure (neighborhoods)?
A: ...*the answer in a few hours*

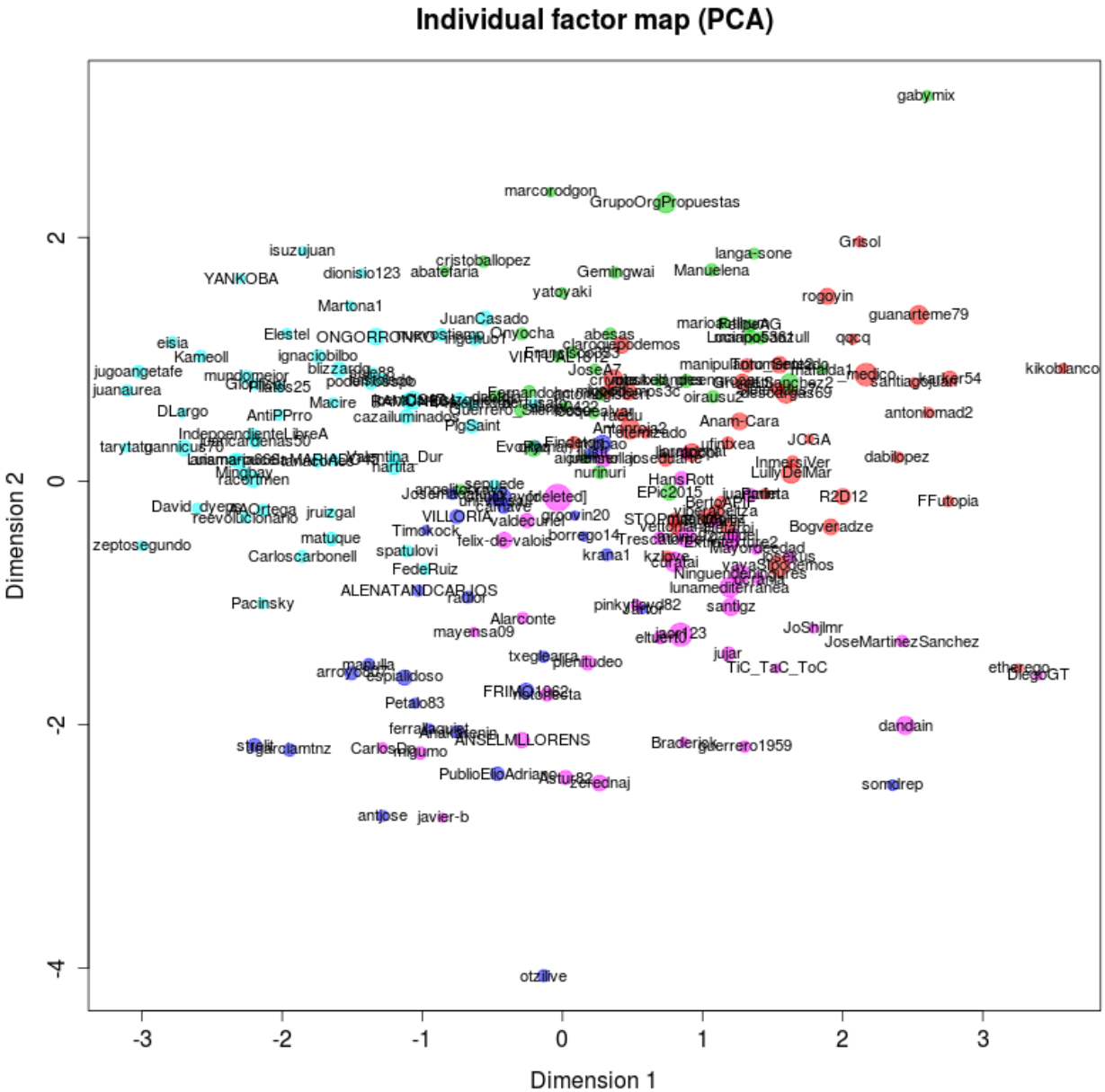
Neighborhood-based clustering

Overview

- Create a user $\times$ neighborhood matrix of counts.
- Z-normalize (users characterized by their deviation from the mean)
- Cluster!

Neighborhood-based clustering

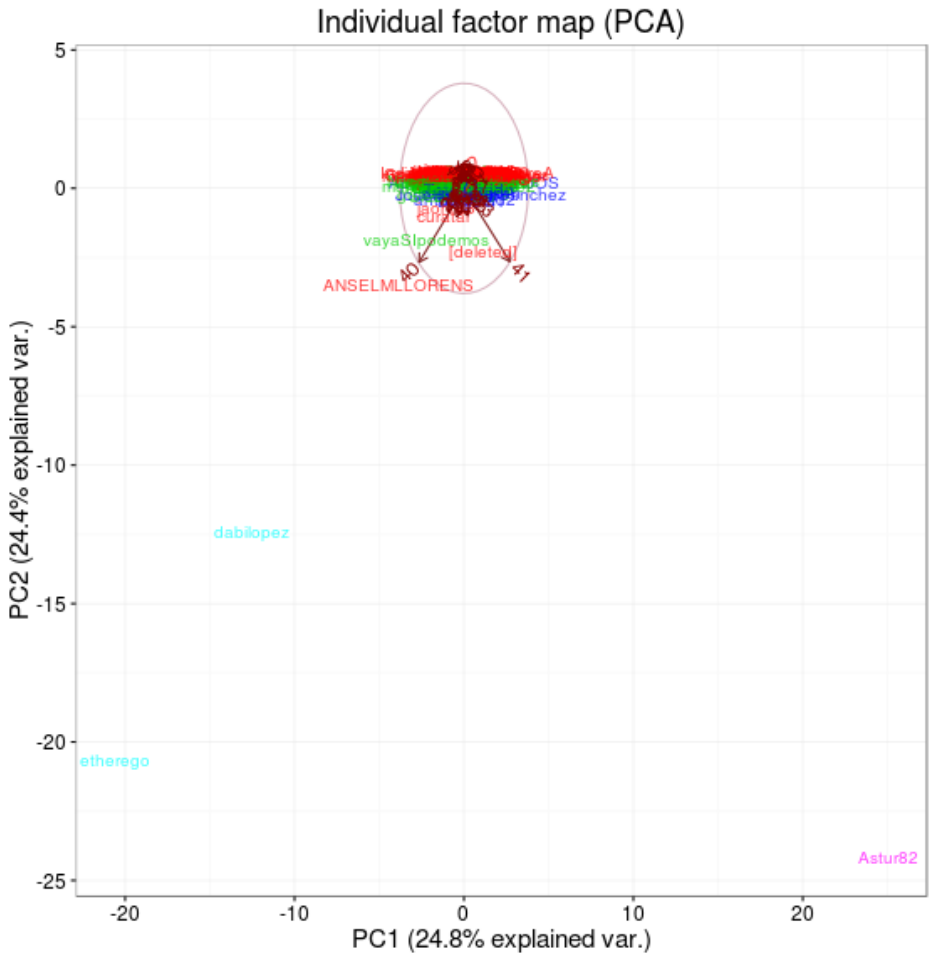
Uncolored neighborhoods



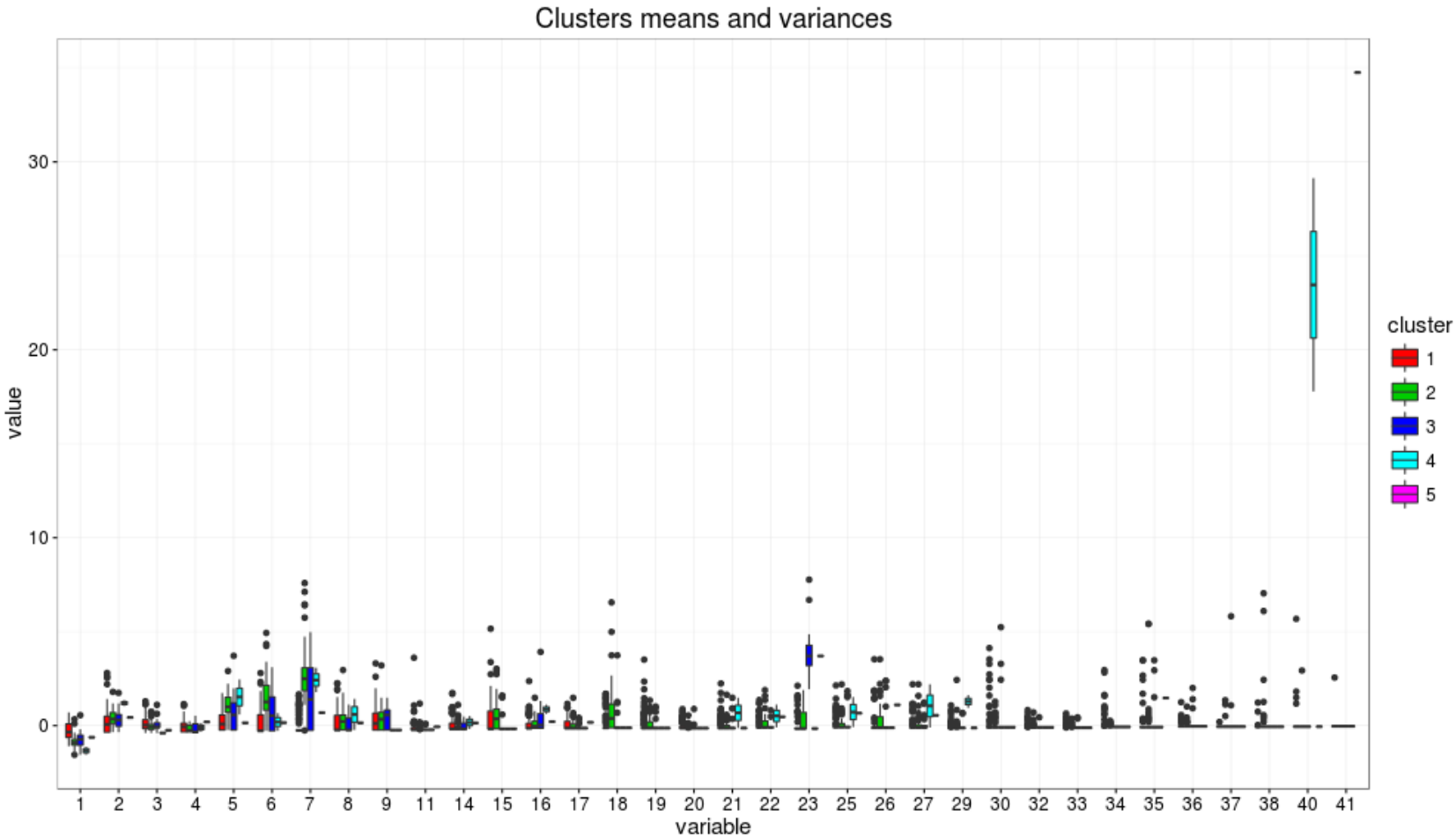
Neighborhood-based clustering

Uncolored neighborhoods

k-means suggests 5 groups:



Neighborhood-based clustering



Conclusions

- Temporal neighborhoods richer than triads to analyze the structure of conversations.
- Users can be characterized in terms of what type of neighborhood they participate in (or they trigger).

Future work:

- Do users jump from cluster to cluster (paths of roles)
- Are initial census predictive of the success of a discussion?.