You are the way you (structurally) talk: Structural-temporal neighbourhoods of posts to characterize users in online forums

Alberto Lumbreras
Jouve B., Velcin J., Guégan, M.

April 8, 2016

Overview

Introduction

- The data
- The graph representations of the data

Structures of conversations

- Basic idea
- Triadic structures
- Neighbourhood structures

Conversation-based clustering

Conclusions

Introduction

- The data
- The graph representations of the data

Structures of conversations

- Basic idea
- Triadic structures
- Neighbourhood structures

Conversation-based clustering

Conclusions

The data

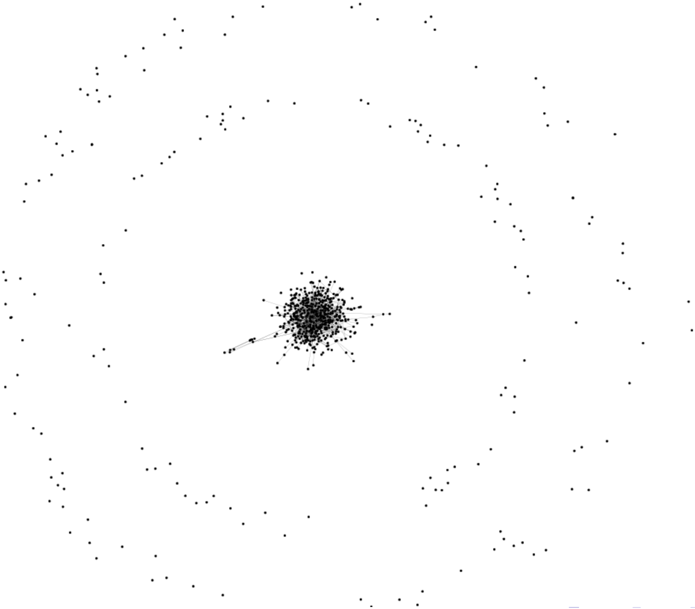
Reddit. A forum of forums

Download monthly dumps 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 ← in this presentation
...

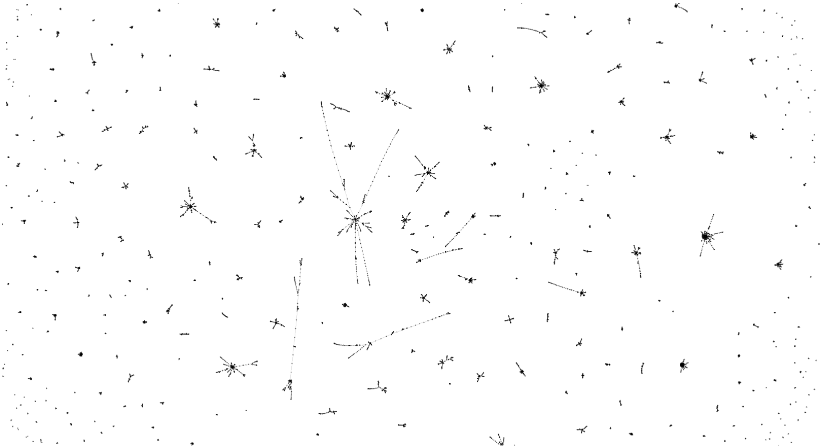
Graph representations

Graph of user interactions (a social network)



Graph representations

Trees of posts



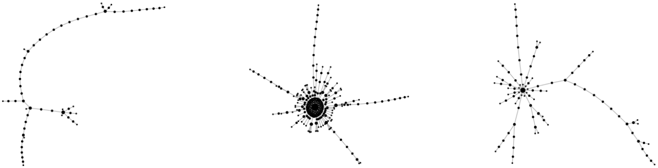
Graph representations

Which one is better?

- Depends on the task!
- One might choose multiple representations (multi-level analysis)

My choice:

- Mostly tree representation
- Because it explicitly represents discussions (and their evolution).



And sometimes:

- SNA representation of single conversations.

Introduction

- The data
- The graph representations of the data

Structures of conversations

- Basic idea
- Triadic structures
- Neighbourhood structures

Conversation-based clustering

Conclusions

Intuition

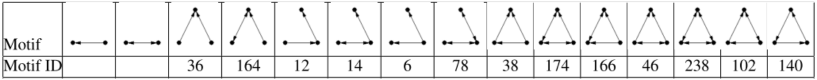
Hypothesis: different individuals have tendency towards different types of conversations and these types are reflected in the structure of their interactions.

These conversational structures might be observed at two levels (at least):

- Social graph.
- Posts graph (tree)

Triadic structures

Triads are not enough



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

Triads in **trees of posts**:

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

Triads in **social graph**:

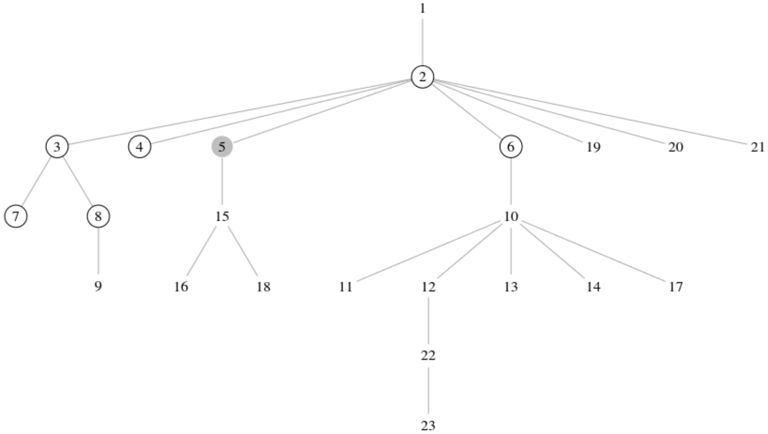
- Order (therefore dynamic) is missing.

We need something richer that captures the dynamics of conversations.

Order-based neighbourhoods

Definition

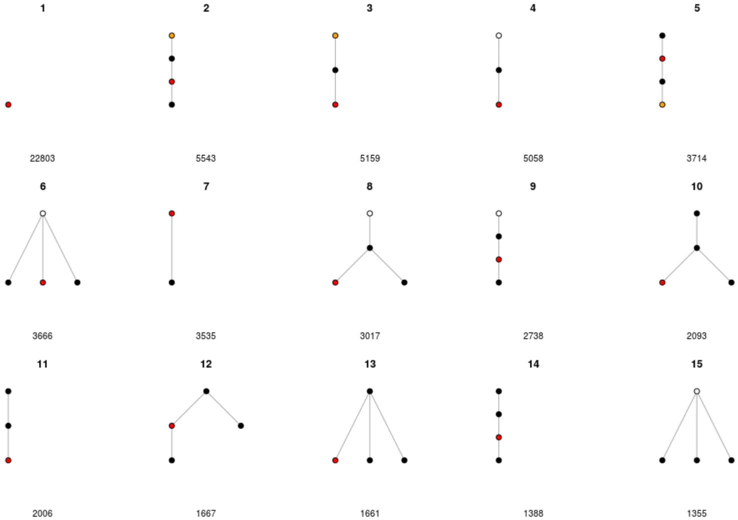
- 1. Extract neighbourhood of post i with radius r.
- 2. Keep only the n posts that are closest (in time) to post i.



Order-based neighbourhoods

Census

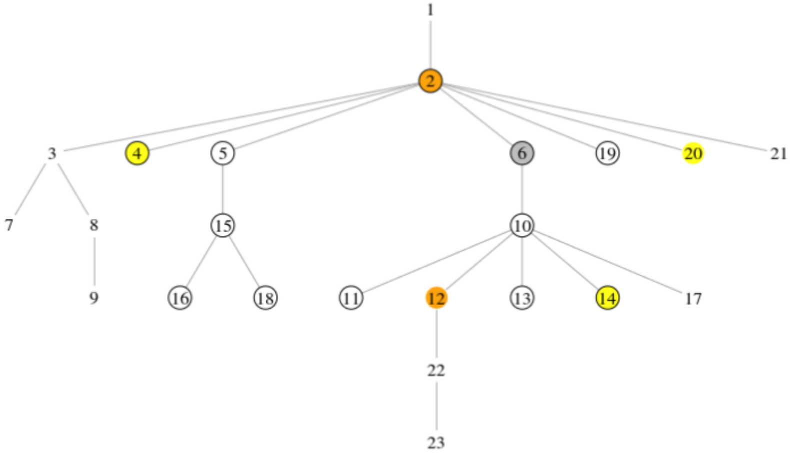
We found 129 different motifs:



Time-based neighbourhoods

Definition

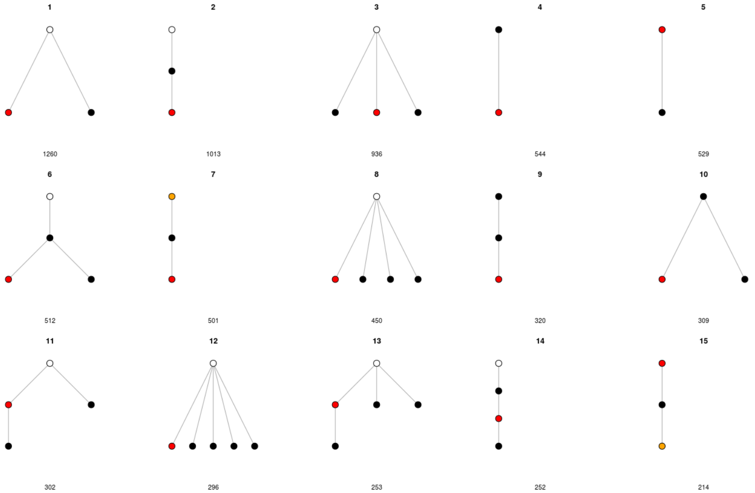
- 1. Extract neighbourhood of post *i* with radius *r*.
- 2. Detect changes of speed (vertical/horizontal changepoints)
- 3. From *i*, get the posts around until a changepoint is found.



Time-based neighbourhoods

Census

We found 165 different motifs:



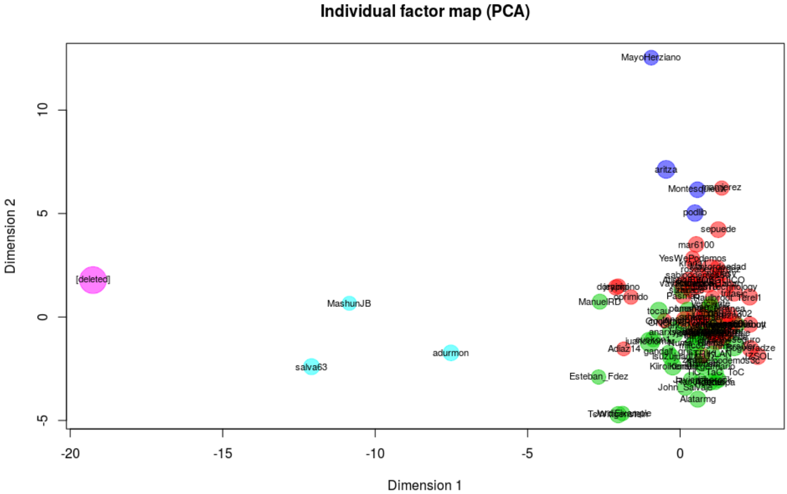
- Introduction
  - The data
  - The graph representations of the data
- Structures of conversations
  - Basic idea
  - Triadic structures
  - Neighbourhood structures
- Conversation-based clustering
- Conclusions

Methodology

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

Conversation-based clustering

Order-based



- Introduction
  - The data
  - The graph representations of the data
- Structures of conversations
  - Basic idea
  - Triadic structures
  - Neighbourhood structures
- Conversation-based clustering
- Conclusions

Conclusions

- **Q: Can we use graph structure to characterise users?**
- A: Yes!

- **Q: By using triads**?
- A: No. They are not useful in trees.

- **Q: So, what kind of structure?**
- A: Posts neighbourhoods that are time/order sensitive.

- **Q: What about language?**
- A: It's ok, but structure is more directly linked to thread dynamics (future work)

Future work:

- Prune time-based neighbourhoods to reduce dimensionality.
- Do users jump from cluster to cluster (paths of roles)

Merci !