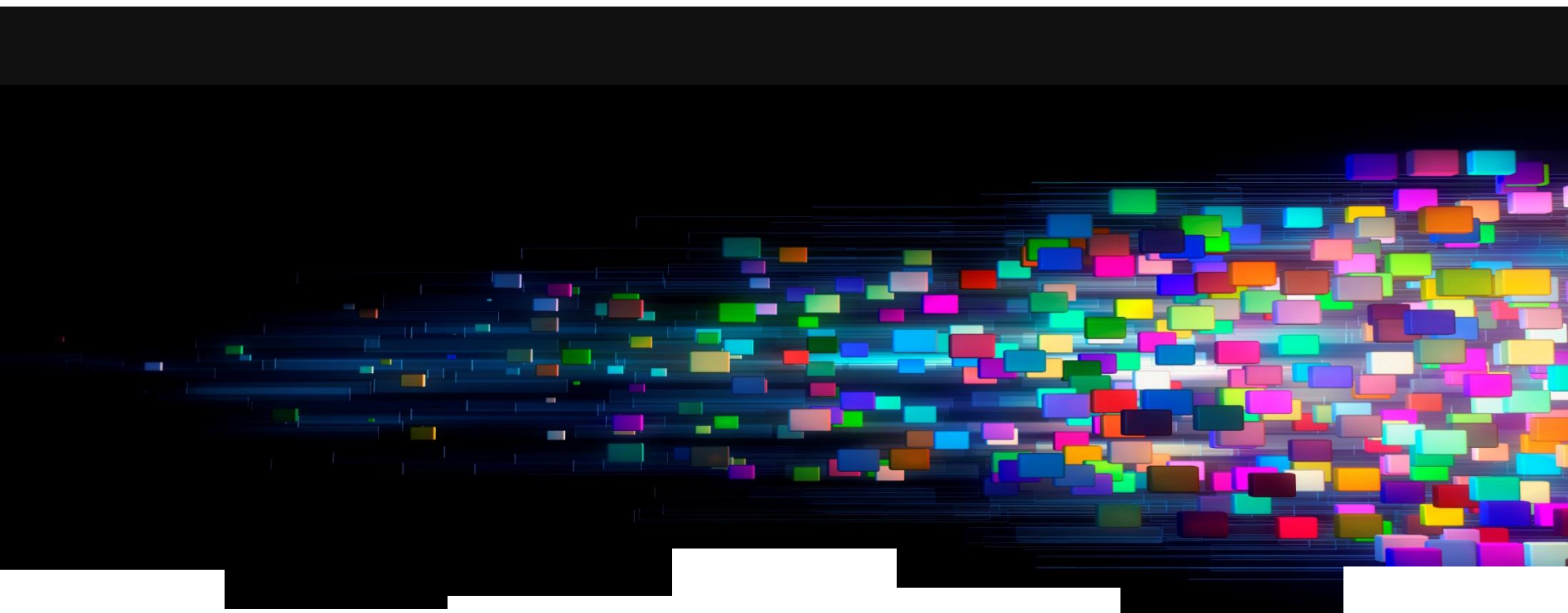


An Insight into the Analysis of Roles in IMDb

Complexity in social systems: from data to models

June 28, 2013



Alberto Lumbreras, James Lanagan (Technicolor)
Bertrand Jouve, Julien Velcin (Université Lyon 2)
Technicolor - Research & Innovation, France



Introduction

Profesor, Doctor, Priest, King.
Mother, Friend, Lover, Celebrity.
Troll, Lurker, Lamer, Flamer, Admin.



WHY ROLES?

- **USERS:**

- To understand (and adapt) the social mechanisms that explain **USERS BEHAVIOR**.

- **(Online) COMMUNITIES:**

- To understand, predict, enhance or prevent changes in the ecosystem of roles that affects the **FITNESS OF A COMMUNITY**.

Outline

- I. Role definition
- II. Approaches to (latent) role detection
- III. First analysis of IMDb
- IV. Conclusions and future work

Outline

- I. Role definition
- II. Approaches to (latent) role detection
- III. First analysis of IMDb
- IV. Conclusions and future work

Structuralism:

- Roles are behavioral expectations that are associated with, and emerge from, identifiable **positions in social structure**.

<<Two entities have the same role if they both are related to other entities in the same pattern.>> (Lorrain)

Structuralism in Computer Science:

- Blockmodels.
- Centrality metrics (e.g.: betweenness).
- Reputation metrics (e.g.: PageRank).
- ...

(Symbolic) Interactionism:

- Study of roles through the social interactions (meanings, etc).

Interactionism in Computer Science:

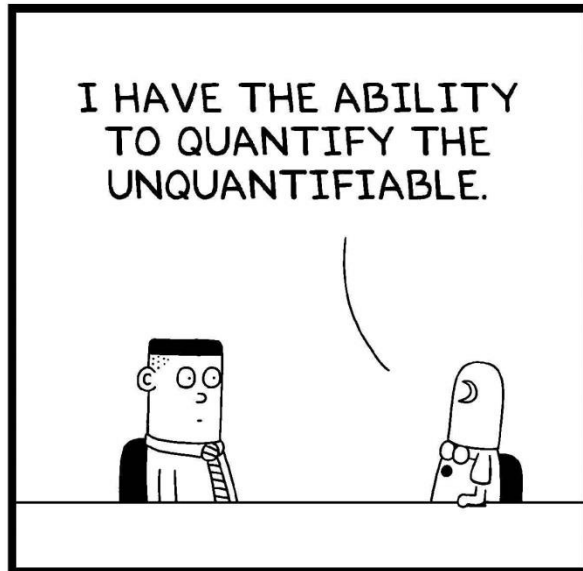
- Natural Language Processing.
- Conversation patterns (e.g.: motifs)

Role definitions in Sociology

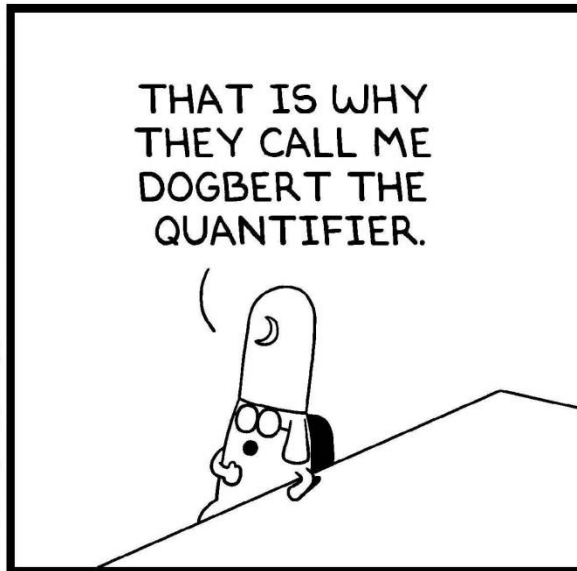
Some specific definitions:

- Roles as **scripts of actors** (Goffman):
 - A script a person plays in front of an audience.
 - Enactment of rights and duties attached to a given status.
- Roles as **social tags** (Skinner):
 - Roles are cognitive concepts used to store social meaning.
 - Social tags that codify how to interact with a tagged person.
- Roles as **expected behaviours** (Dictionary of Psychology)
 - The set of behaviours expected of a person possessing a certain social status.

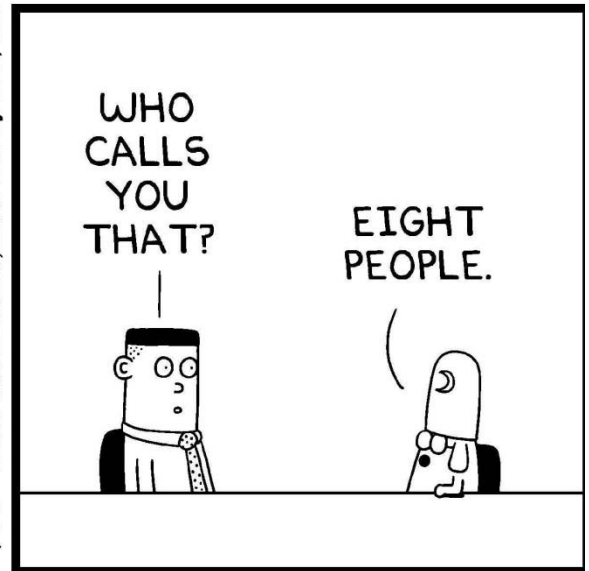
Quantifying roles



www.dilbert.com
scottadams@aol.com



4-5-07 © 2007 Scott Adams, Inc./Dist. by UFS, Inc.



Quantifying roles

Some specific definitions:

- Roles as scripts of actors (Goffman):
 - A **SCRIPT** a person plays in front of an audience.
 - Enactment of **RIGHTS** and **DUTIES** attached to a given status.
- Roles as social tags (Skinner):
 - Roles are **COGNITIVE** concepts used to store **SOCIAL MEANING**.
 - Social tags that codify how to **INTERACT** with a tagged person.
- Roles as expected behaviours (Dictionary of Psychology)
 - The set of **BEHAVIORS** expected of a person possessing a certain **SOCIAL STATUS**.



Measurable



Non measurable

Quantifying roles

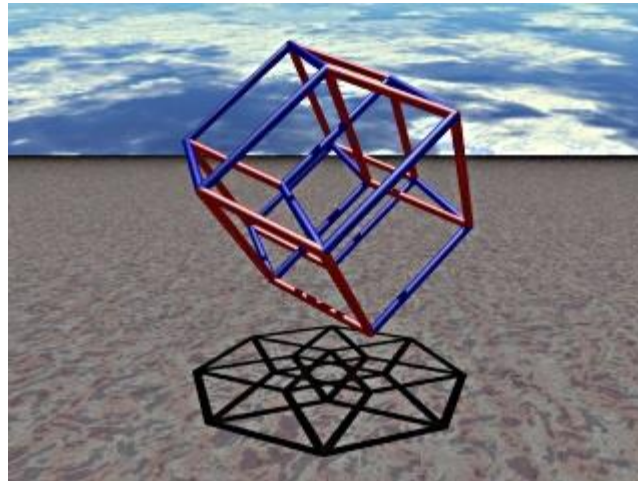
*[The intuition is that] Any role is a combination of particular sets of **BEHAVIORAL**, **MEANINGFUL**, and **STRUCTURAL** attributes.” (Welser)*

(user)

(community)

BEHAVIOR, MEANING, STRUCTURE are
observable projections of roles.

(for a machine [for now])



Outline

- I. Role definition
- II. Approaches to (latent) role detection
- III. First analysis of IMDb
- IV. Conclusions
- V. Discussion and Future work

Role Detection: approaches

Top Down

(from roles to data)

- Find roles (defined a priori) in a community
- Techniques:
 - Supervised learning (?)
 - Metrics analysis
 - Visualization

Bottom Up

(from data to roles)

- Hypothesis: set of relevant features or phenomena.
- Techniques:
 - Unsupervised learning
 - Blockmodels
 - Visualization

Ethnography

- Requires to «live» in the community.
- Captures meaning, expectation...
- The most complete, the less automatizable.



Role Detection: approaches



- Roles emerge from **communities**.



- No silver bullet
 - Communities nature can change from forum to forum.
 - Universal roles (e.g.: troll)
 - Forum-specific roles (e.g.: expert)



- **Know your community!**
 - Are there communities in the forum?
 - How are them? (structure, dynamics, norms...)

Outline

- I. Role definition
- II. Approaches to (latent) role detection
- III. **First analysis of IMDb**
- IV. Conclusions
- V. Discussion and Future work

Analysis of IMDb - Internet Movie Database

- Discussion threads (topics) about films content
- Our dataset
 - Films sample:
 - Top 250 + Top 1000 with at least to 7500 votes + Bottom 100
 - 706 films
 - 1M posts
 - 102K threads
 - 117K users (at least 1 post)



IMDb Find Movies, TV shows, Celebrities and more... All

Movies TV News Showtimes Community IMDbPro Apps

Board: Le couperet (2005) [My Profile](#) | [Main Boards](#) | [Fri](#)

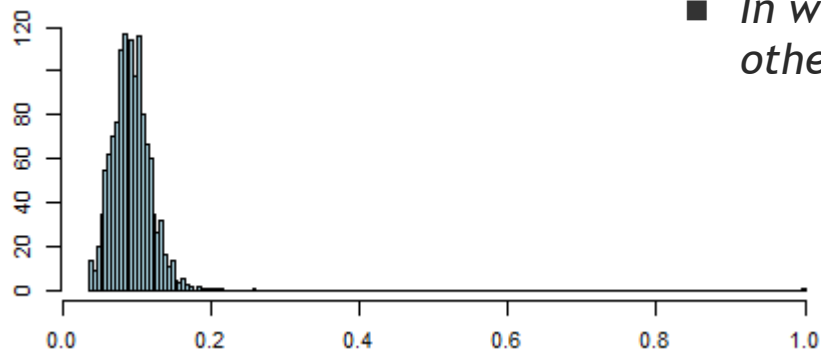
This board is reserved for discussions about: [Le couperet \(2005\)](#)
new Take the [Le couperet \(2005\) Quiz!](#)

Inappropriate discussions may be deleted or moved

[Start New Topic](#) [Add to Favorites](#)

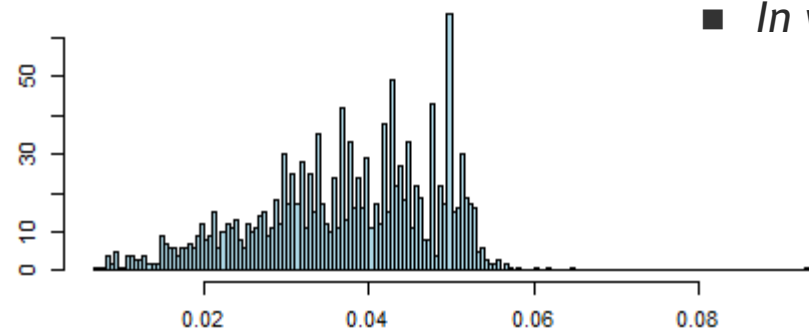
Subject	Started by	Replies	Latest Posting
Genre?	nisitpav	7	Thu Feb 7 2013 06:15
Anyone reminded of Jack Lemmon?	KennedyMarbella	4	Sat Feb 2 2013 18:26
many doubts....SOME SPOILERS	natakalt-1	2	Wed Jul 6 2011 01:52
Things we have learned from this movie	is-csi	4	Tue Feb 8 2011 14:57
'This is for the next customer'..?	aznfx	1	Tue Feb 8 2011 14:39
The End?	perhapsperpetua	23	Thu Nov 25 2010 03:43
What's up with all the poster ads?	frank_h2	7	Thu Sep 2 2010 13:40
Chan-Wook Park to remake this movie	sunn-o	0	Mon Mar 22 2010 09:07
What the cop says when he leaves?	mtherfackr	1	Tue Oct 27 2009 07:12
[Post deleted]	Deleted	1	Thu Sep 10 2009 09:57
North American Release	Bellzy	3	Mon Dec 31 2007 19:08
DVD Release	catabatic-1	3	Mon Sep 24 2007 12:55

Analysis of IMDb - friendship analysis



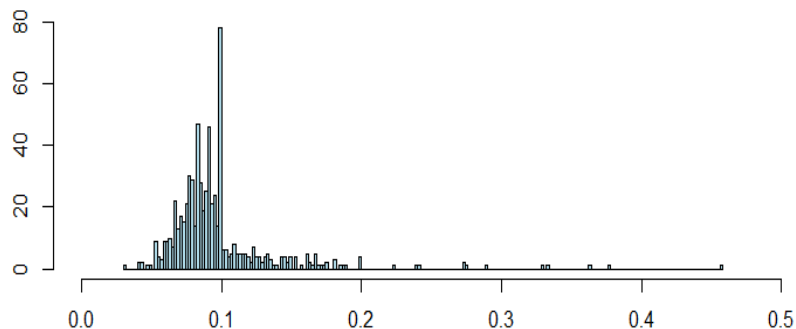
- *In what % of films does a given user co-participate with other user?*

$$coincidences_u(v) = \frac{common_films(u, v)}{films(u)}$$



- *In what % of films do user u interact with an old friend?*

$$interactions_u(v) = \frac{replies(u, v)}{films(u)}$$



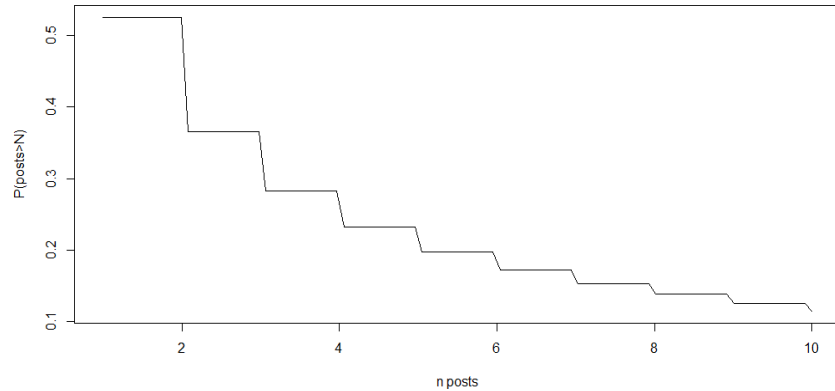
- *Given friends u and v co-participate in the same film, what is the probability of an interaction? (more than 10 coincidences)*

$$conditional_interaction_u(v) = \frac{replies(u, v)}{coincidences_u(v)}$$

Analysis of IMDb - statistics of participations

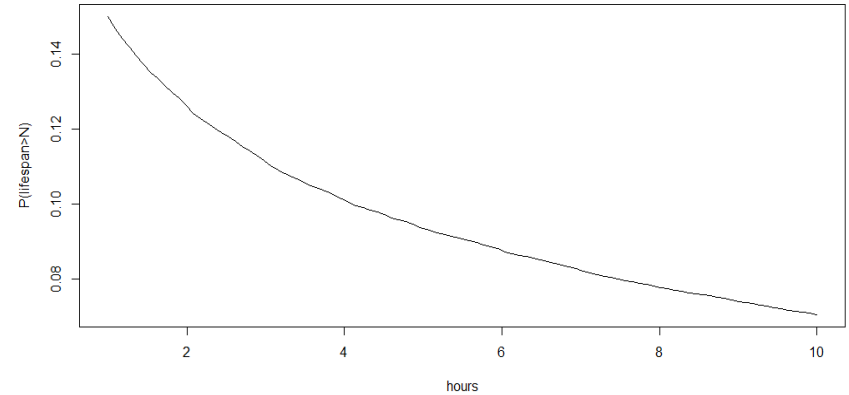
■ *Posts/user*

Total posts per user (survival)



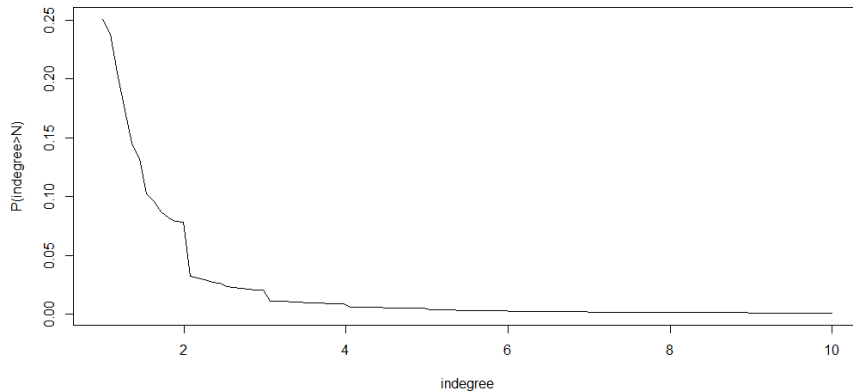
■ *Life/user in a thread*

Users lifespan per thread (survival)

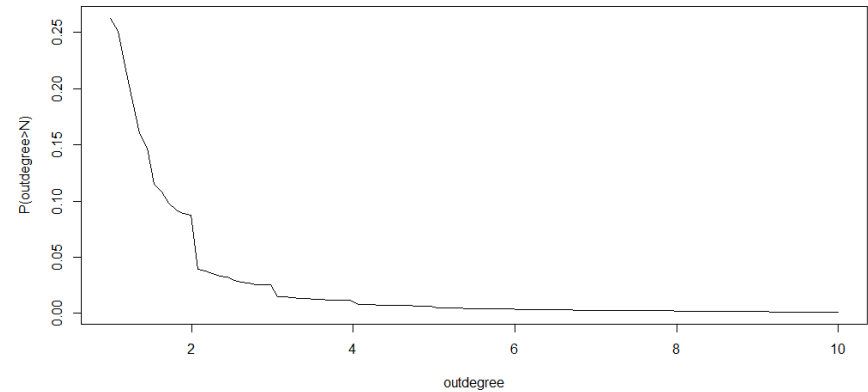


■ *Degrees/user in a thread*

Average users indegree per thread (survival)



Average Users outdegree per thread (survival)



SEGMENTED ANALYSIS

Analysis of IMDb - segmented analysis

1-posters

- Wrote 1 post and then lurked or abandoned
 - 47.5% of users

Shy

- Participated in 1 thread and then lurked or abandoned
 - 5.3% of users

Regular users

- Between 2 and 10 threads on IMDb
 - 37% of users

Pros

- Between 10 and 100 threads on IMDb
 - 9% of users

Super-actives

- More than 100 threads on IMDb
 - 0.7% of users

Analysis of IMDb - segmented analysis

	Films	Threads	Posts	Words/post	Thread length	Posts/thread	In/thread	Out/thread	Life	In/out	Birth
1-poster	1	1	1	36	31	1	1	1	0	0	55
Shy	1	1	2	50	31	2	2	2	2.8	1	1.7
Regular	2	3	3	46	31	1	0.5	1	0	1.2	106
Pro	9	18	24	52	29	1.4	1.23	1.27	0.84	150k	107
Super	15	160	274	50	25	1.53	0.78	1.34	2.6	76k	24

- Superactives impulsive (and not really social)?
- Shy are more engaged in a thread. (they start 50% of their threads)
- Short lives in threads.

Analysis of IMDb - segmented analysis

- How often do you start a thread?
 - One-posters: 16%
 - Shy: 41%
 - Regular: 18%
 - Pros: 16%
 - Superactives: 11%

- And when you start a thread, how often do they ignore you?
 - One-posters: 78% (else if thread already started: 48%)
 - Shy: 35% (else 37%)
 - Regular: 61% (else 41%%)
 - Pros: 57% (else 37%)
 - Superactives: 50% (else 36%)

- One-posters highly ignored (it might take time)
- Shy are thread starters
- If you want to meet people, engage yourself in a discussion

PATTERNS OF CONVERSATIONS

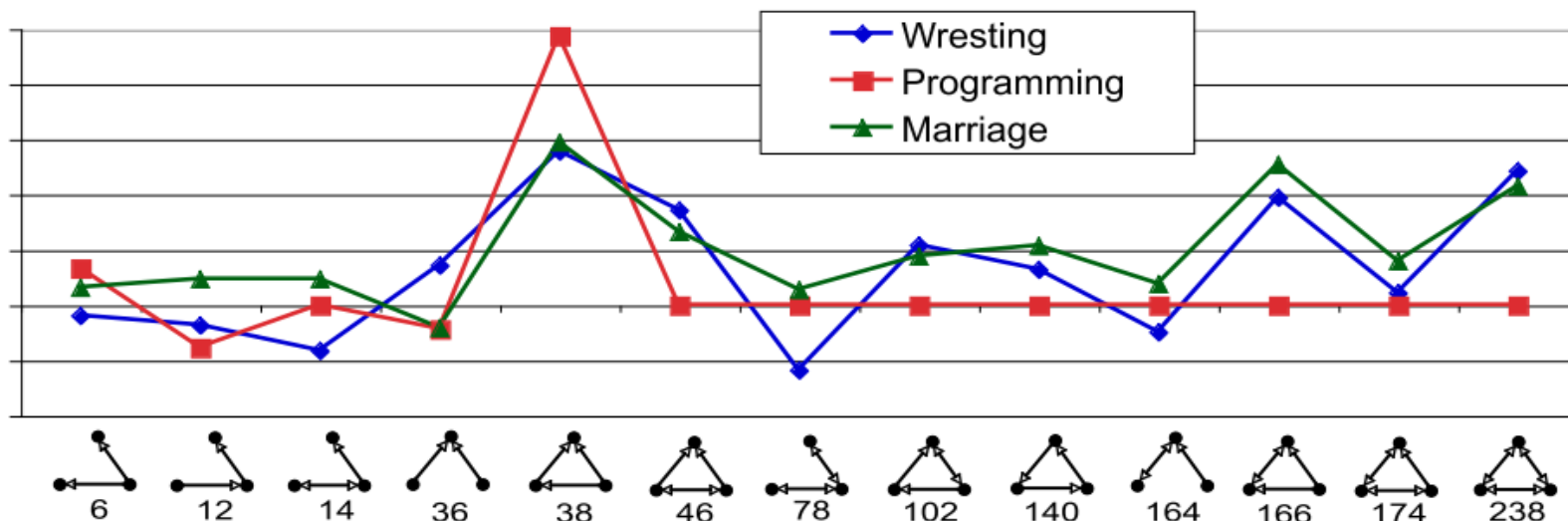
Analysis of IMDb - Patterns of conversations

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

↑
star
↑
cascade
↑
loop

E.g:

Q&A forums are populated by stars
Flame wars are populated by cascades



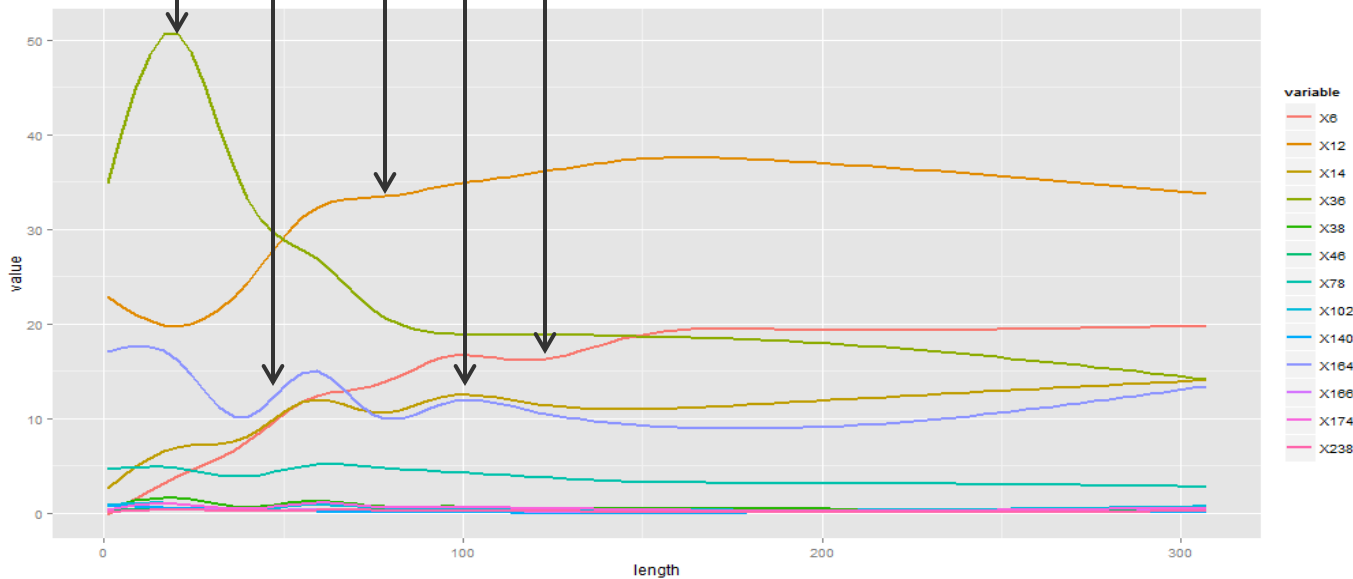
L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman, "Knowledge sharing and yahoo answers,"
in *WWW'08*

technicolor



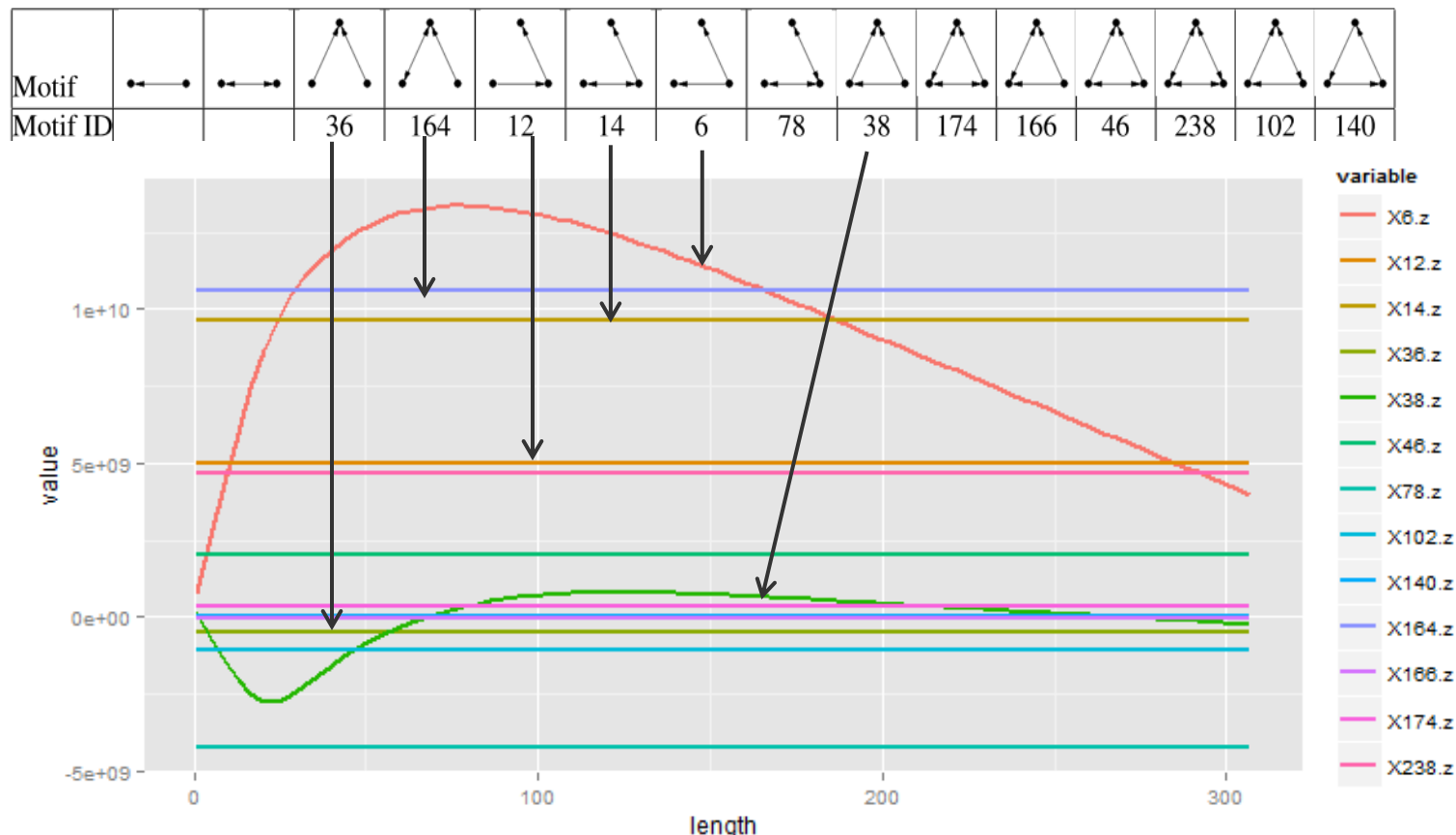
Analysis of IMDb - Patterns of conversations (%)

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



- Short threads: Q&A pattern (36)
- Long threads: cascade pattern (12)
- Question user conversates with some of her repliers (164)
- Some conversations between users (not the seed) in long threads (14)

Analysis of IMDb - Patterns of conversations (z-score)



- Short threads: Q&A pattern (36)
- Long threads: cascade pattern (12)
- Question user conversates with some of her repliers (164)
- Some conversations between users (not the seed) in long threads (14)

Outline

- I. Role definition
- II. Approaches to (latent) role detection
- III. First analysis of IMDb
- IV. Conclusions and future work

Conclusions

■ ROLES

- There is no unified definition. (and there might never be)
- ... but general agreement on common elements:
 - behavior,
 - structure,
 - meaning.

■ IMDb

- Conversations:
 - Conversations ad-hoc (users don't know each other from other films (threads?))
 - Different conversations in short and long threads.
- Communities:
 - No evidence of communities in IMDb (yet: microcommunities?)
 - Social features could improve participation. (e.g.: integration with other SN)

■ Conversation patterns (motifs)

- Comprehensible technique to understand conversations.



Future work

■ ROLES

- Role detection must include time.
(behavior, structure and even meaning are dynamic concepts)

■ IMDb (and forums)

- Analysis of evolution of shy, regulars, pros, and super actives.
 - How did they evolve from « 1 posters » and <<shy>>?
- Can the initial population of motifs predict the length of a thread?
- Better crawling strategy.
- Benchmark with other forums.

■ Conversation patterns (motifs)

- Analysis with temporal motifs.
- Do different roles relate to different motifs?
 - Motifs as users' features for role detection.

■ A mix bottom up - top down strategy for role detection

MERCI!
