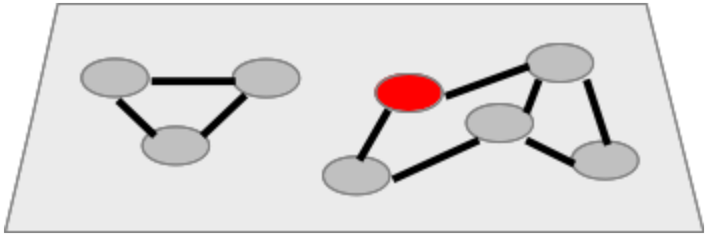
UPPSALA UNIVERSITET

Moving out of flat-land

analysis and mining of multiple social networks

Matteo Magnani
Uppsala university
matteo.magnani@it.uu.se

Social Network Analysis



Activity slide

1) Work individually.
2) Grab paper and pen (or equivalent technology).
3) Think of the people constituting your research group. (between 7 and 12 people?)
4) Draw the network of your working relationships (= who you are directly collaborating with).

You have 3 minutes.

Activity slide

1) Work individually.
2) Take your working network.
3) Use a different color / line type and add your *fika* network on top of it.
4) Use a different color / line type and add your *friendship* network on top of it.
5) Use a different color / line type and add your *facebook* network on top of it.

You have 3 minutes.

Some traditional questions, revisited

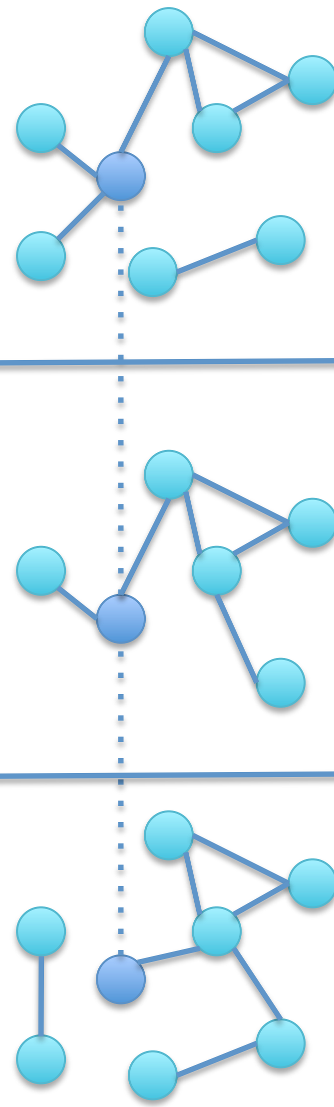
- Which individuals should know about you and the fact you are a brilliant student?
- How far are you from that PhD student you would like to invite out for dinner?
- Are there any research sub-groups you might want to join?

From user-centered to dimension-centered

Which dimensions determine the popularity of a user?

Which dimensions keep two users close to each other?

Which dimensions define a user’s communities?



What is the popularity of a user?

How close are two users?

In which community can we classify a user?

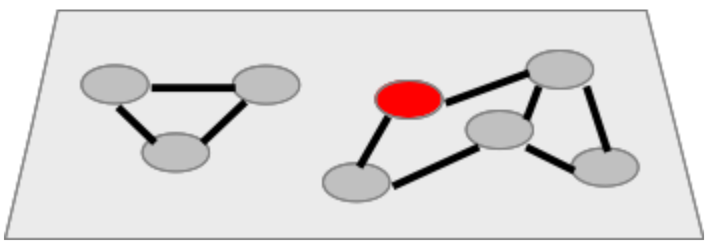
UPPSALA UNIVERSITET

Larger-scale questions

- Government initiatives to shut down Twitter.
- Etc. etc. etc.

One-minute-break slide

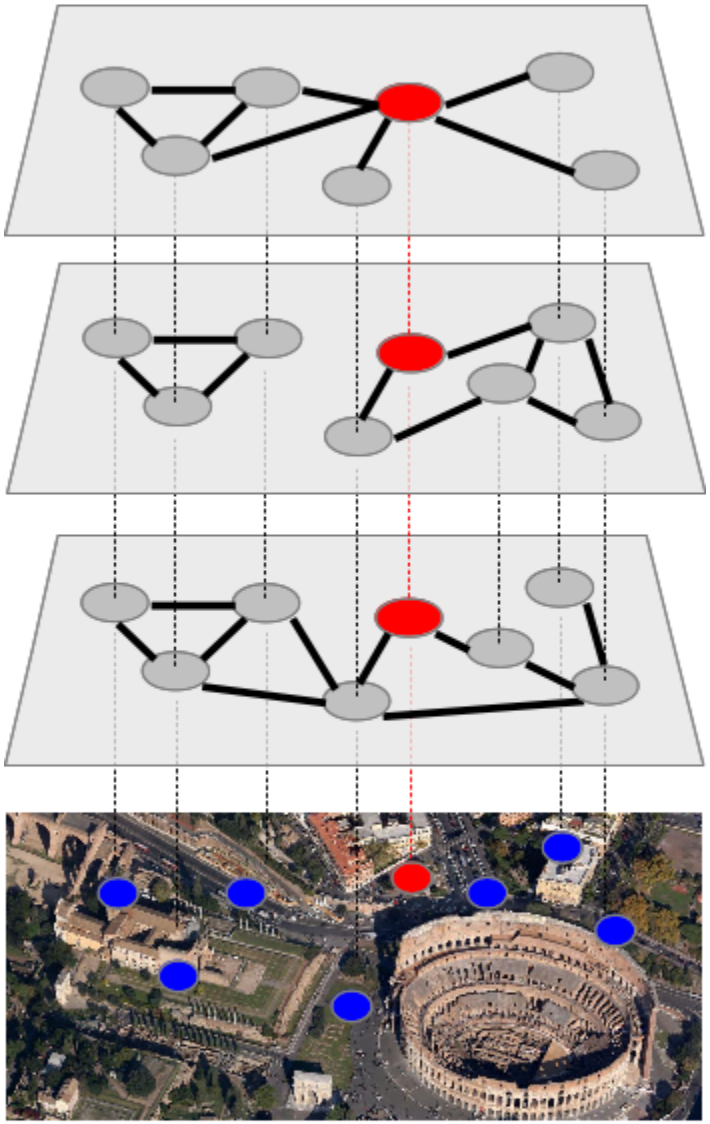
Social Network Analysis



twitter

UPPSALA UNIVERSITET

Multiple Social Network Analysis



friendfeed

YouTube

twitter

off-line encounters

UPPSALA UNIVERSITET

Why this tutorial?

Multiple networks

Multiplex networks

Networks of networks

Multi-modal networks

Multidimensional networks

Multilevel networks

Multi-layer networks

Labeled graphs

Heterogeneous information networks

Plan

- 9:00-9:25 - Introduction
- 9:25-10:00
  - Part I: Historical foundations and models. | Background
- 10:00-10:30
  - Coffee break.
- 10:30-12:30
  - Part II: Measures. | SNA
  - Part III: Formation. | Dynamics
  - Part IV: Community detection (if time left). | Mining
  - Part V: Discussion.


Understanding social processes might require the analysis of multiple interdependent networks.

UPPSALA UNIVERSITET

- Multiplexity as a quality measure
- Relation specific approach

Similarities			Social Relations				Interactions	Flows
Location	**Membership**	**Attribute**	**Kinship**	**Other role**	**Affective**	**Cognitive**	e.g.,	e.g.,
e.g., Same spatial and temporal space	e.g., Same clubs Same events etc.	e.g., Same gender Same attitude etc.	e.g., Mother of Sibling of	e.g., Friend of Boss of Student of Competitor of	e.g., Likes Hates etc.	e.g., Knows Knows about Sees as happy etc.	Sex with Talked to Advice to Helped Harmed etc.	Information Beliefs Personnel Resources etc.

Borgatti et al. 2009 - Network Analysis in the Social Sciences

UPPSALA UNIVERSITET

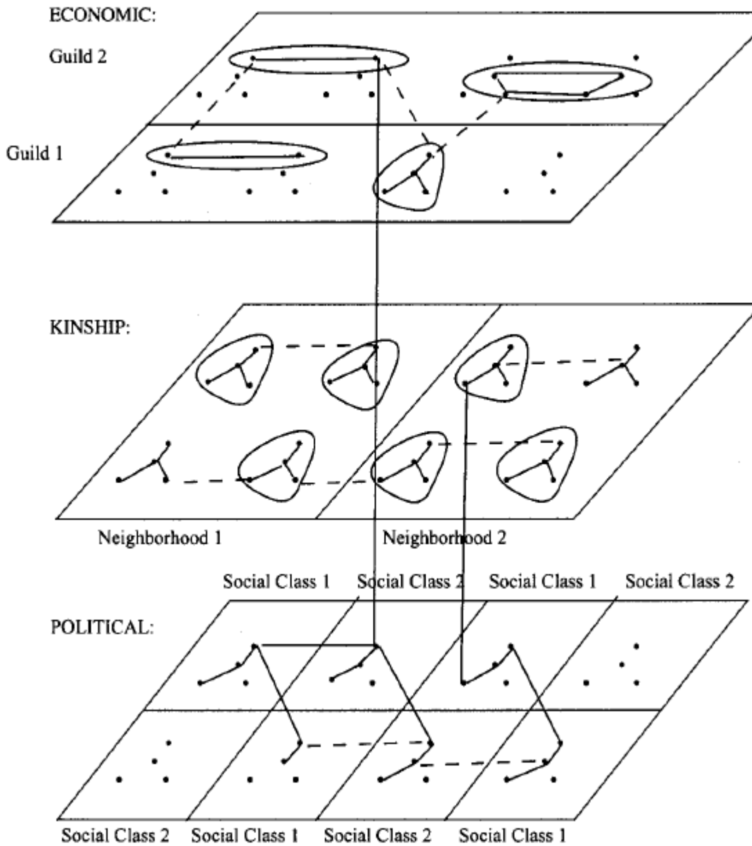


FIG. 1.—Multiple-network ensemble Renaissance Florence. Solid lines are constitutive ties, dotted lines are relational social exchanges, and oblongs are formal organizations (families and firms). People in multiple roles are vertical lines connecting corresponding dots in the domains of activity in which people are active (only two are shown for illustration).

J F Padgett & P D McLean (2006) Organizational Invention and Elite Transformation: The Bir
Partnership Systems in Renaissance Florence. American Journal of Sociology Volume 111 Nu

Social Networks 10 (1988) 383–411
North-Holland

NETWORK MODELS: SOME COMMENTS ON PAPERS IN THIS SPECIAL ISSUE

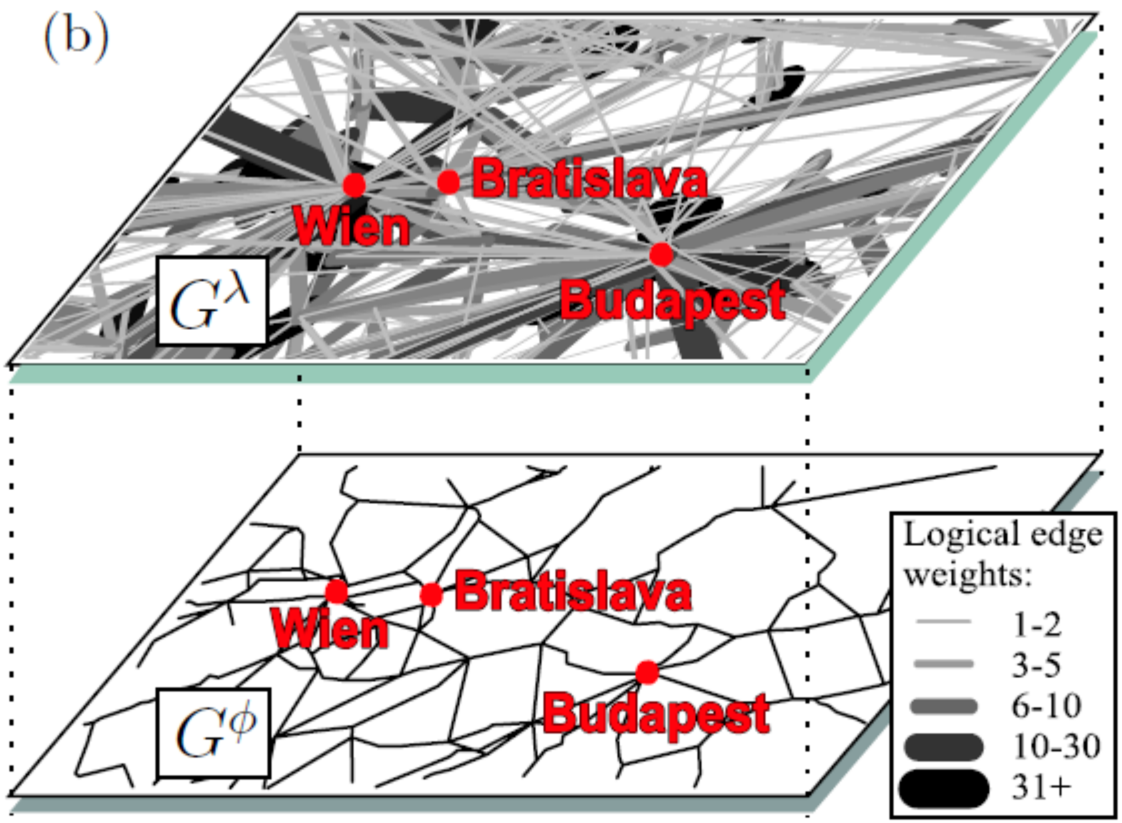
Philippa E. PATTISON *
University of Melbourne

4. Models for network interrelations

Much work in developing representations for multiple networks has focussed on the consequences of different models for describing persons or positions and their interrelations in a network. The papers by

White, H.C., S.A. Boorman and R.L. Breiger
(1976) "Social structure from multiple networks: I. Blockmodels of roles and positions". *American Journal of Sociology 81*: 730–780.

Interdependent networks



Kurant, M., & Thiran, P. (2006). Layered Complex Networks. Physical Review Letters, 96(13), 138701.

Networks of networks

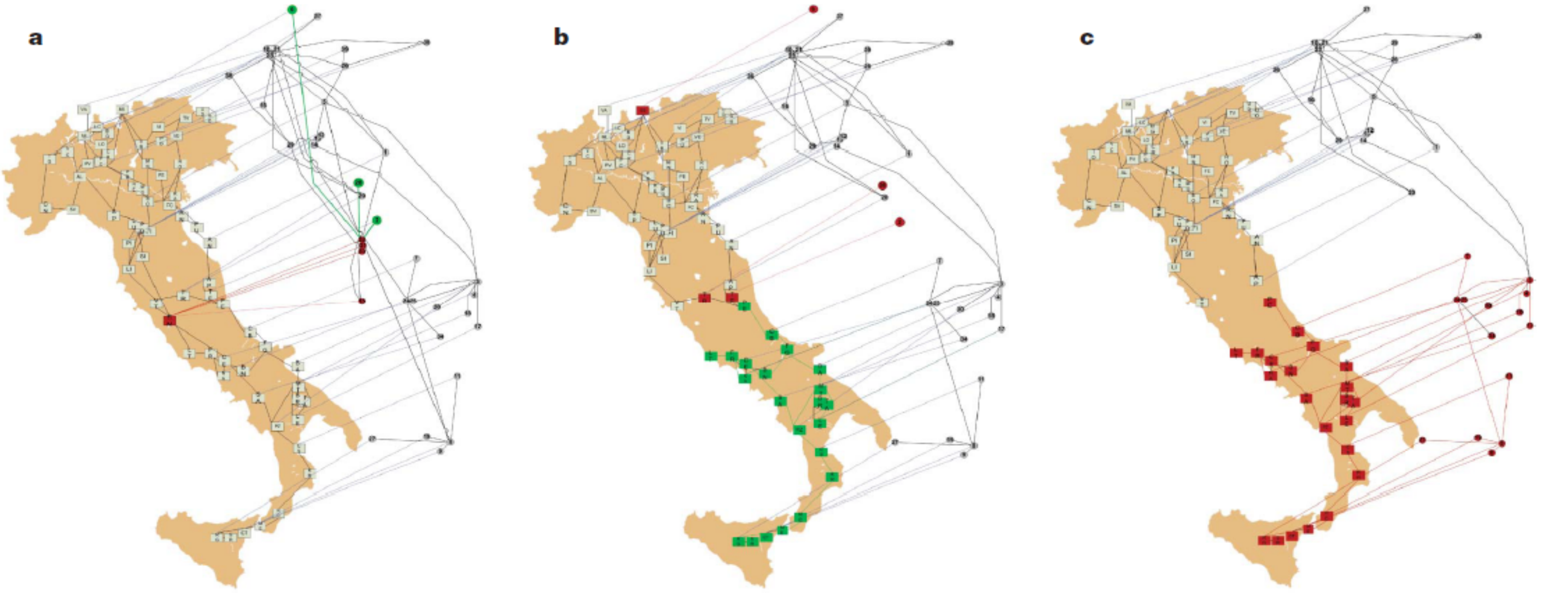
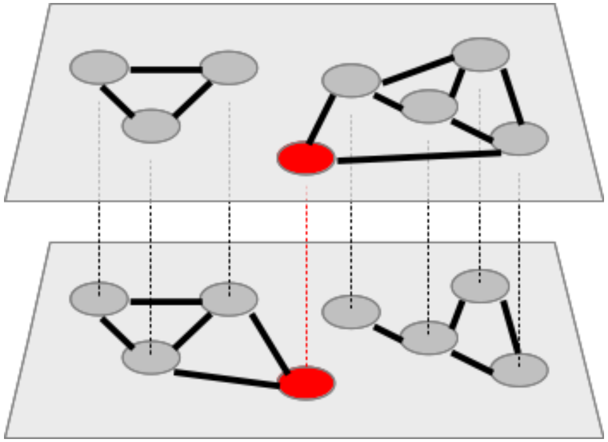
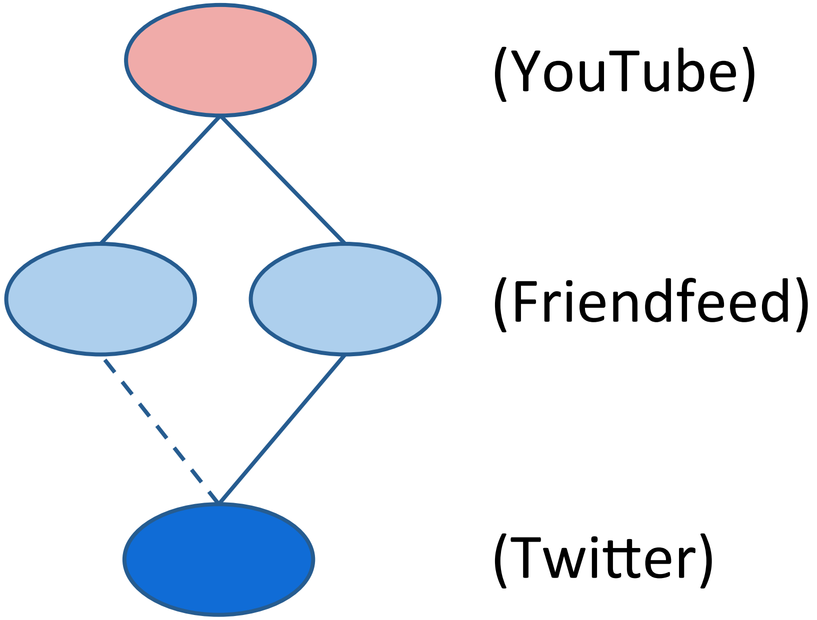


Figure 1 | Modelling a blackout in Italy. Illustration of an iterative process of a cascade of failures using real-world data from a power network (located on the map of Italy) and an Internet network (shifted above the map) that were implicated in an electrical blackout that occurred in Italy in September 2003[20]. The networks are drawn using the real geographical locations and every Internet server is connected to the geographically nearest power station. **a,** One power station is removed (red node on map) from the power network and as a result the Internet nodes depending on it are removed from the Internet network (red nodes above the map). The nodes that will be disconnected from the giant cluster (a cluster that spans the entire network) at the next step are marked in green. **b,** Additional nodes that were disconnected from the Internet communication network giant component are removed (red nodes above map). As a result the power stations depending on them are removed from the power network (red nodes on map). Again, the nodes that will be disconnected from the giant cluster at the next step are marked in green. **c,** Additional nodes that were disconnected from the giant component of the power network are removed (red nodes on map) as well as the nodes in the Internet network that depend on them (red nodes above map).

Buldyrev, S. V, Parshani, R., Paul, G., Stanley, H. E., & Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. Nature, 464(7291), 1025–8.

Multiple social networks (with complex ties)





Magnani M and Rossi L (2011) *The ML-model for multi layer network analysis*. In: ASONAM Conference, IEEE Computer Society.

"Not then, men and their moments. Rather, moments and their men"

Goffmann 1967 - Interaction rituals

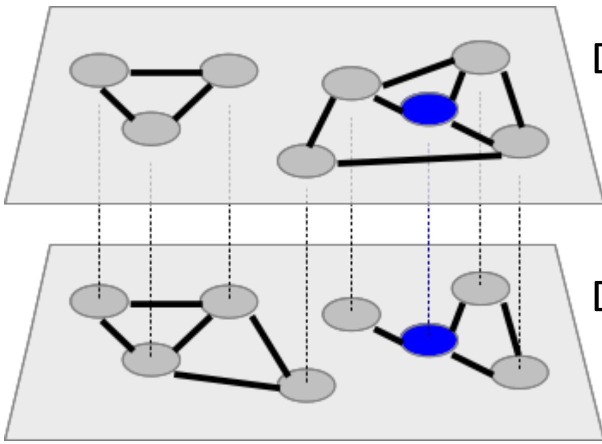
Wrap up

- Multiple network analysis and mining allows us to work on more accurate representations of the world.
- Some questions need this level of detail.
- Different models are "formally similar" but emphasize different aspects of the networks.

UPPSALA UNIVERSITET

NODE MEASURES

Degree and neighborhood

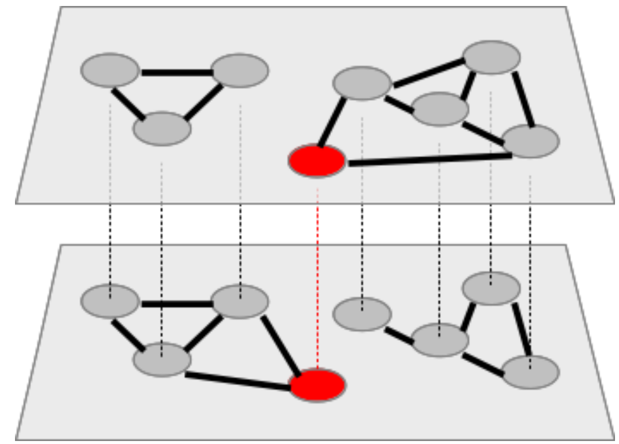


Degree: 3
Central

Degree: 3
Central

Degree: 2
Peripheral

Degree: 2
Peripheral

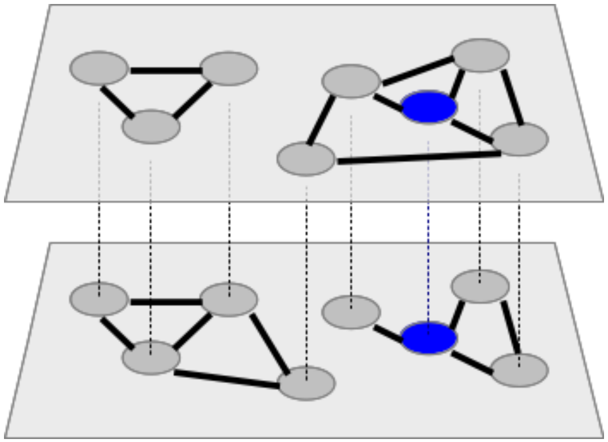


Neighbors: 3
Central

Neighbors: 4
Very central

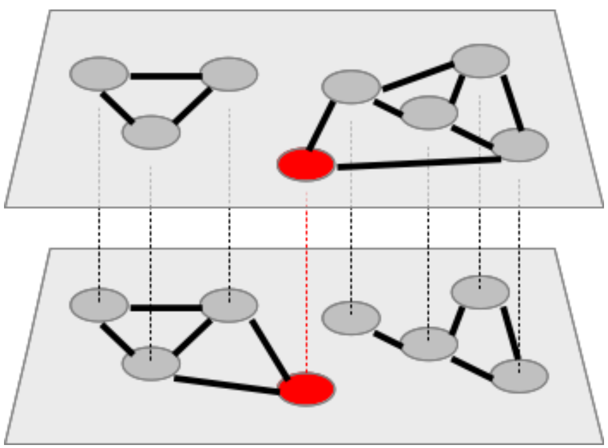
Magnani M and Rossi L (2011) *The ML-model for multi layer network analysis*. In: ASONAM Conference, IEEE Computer Society.

Network relevance



3 neighbors: **relevance 100%**

3 neighbors



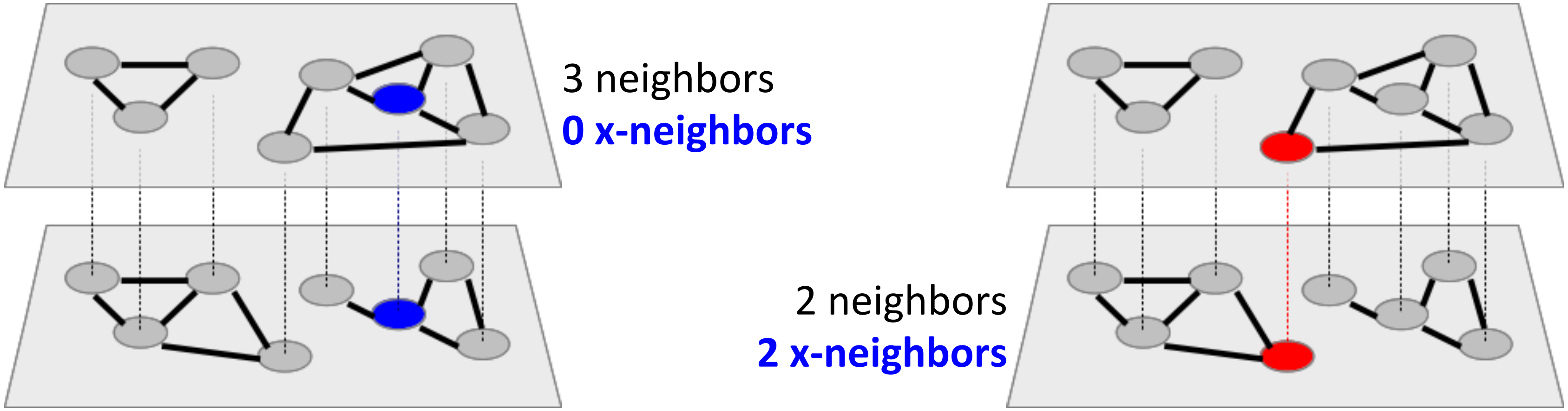
2 neighbors: **relevance 50%**

4 neighbors

$$DimRelevance(v, D) = \frac{Neighbors(v, D)}{Neighbors(v, L)}$$

Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A., & Pedreschi, D. (2012). Multidimensional networks: foundations of structural analysis. WWW Journal

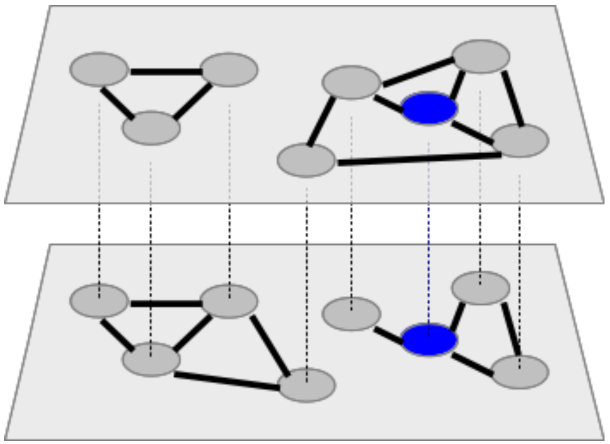
Exclusive neighbors



$$Neighbors_{XOR}(v, D) =$$

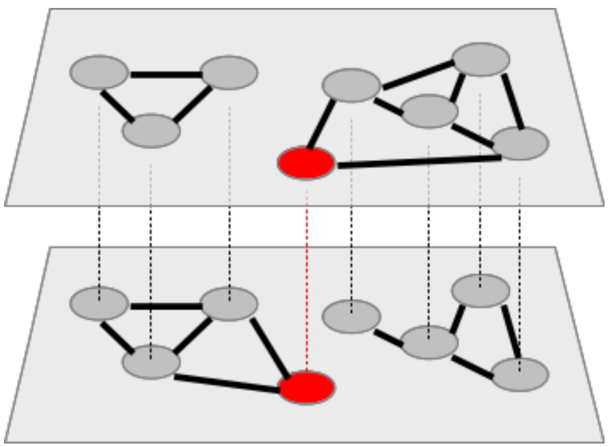
$$\left|\{u \in V \mid \underbrace{\exists d \in D : (u,v,d) \in E}_{} \wedge \underbrace{\nexists d' \notin D : (u,v,d') \in E}_{}\}\right|$$

Exclusive network relevance



3 neighbors

3 neighbors: relevance 100%
0 x-neighbors
x-relevance: 0%



4 neighbors

2 neighbors: relevance 50%
2 x-neighbors
x-relevance: 50%

$$DimRelevance_{XOR}(v, D) = \frac{Neighbors_{XOR}(v, D)}{Neighbors(v, L)}$$

UPPSALA UNIVERSITET

MEASURES: EXAMPLES

UPPSALA UNIVERSITET

AUCS Dataset

- 61 employees of a University department
- Survey-based/Automatic data collection
- 5 kinds of relationships:
  - Coworking,
  - Coauthorship,
  - Friendship (having fun together),
  - Facebook friendship,
  - Having lunch together.

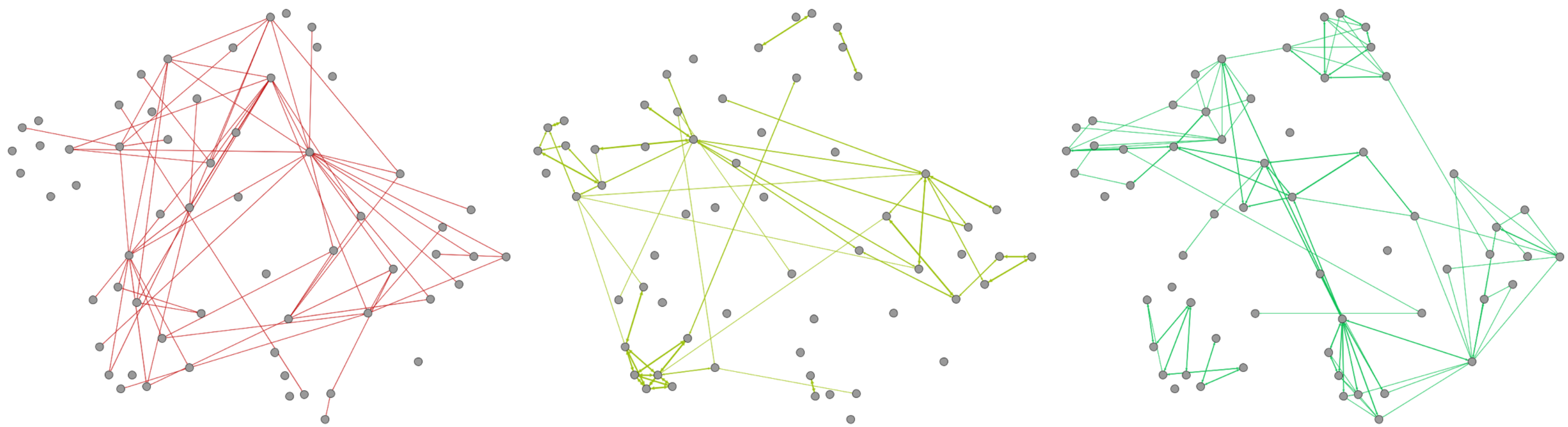
UPPSALA UNIVERSITET

AUCS Dataset

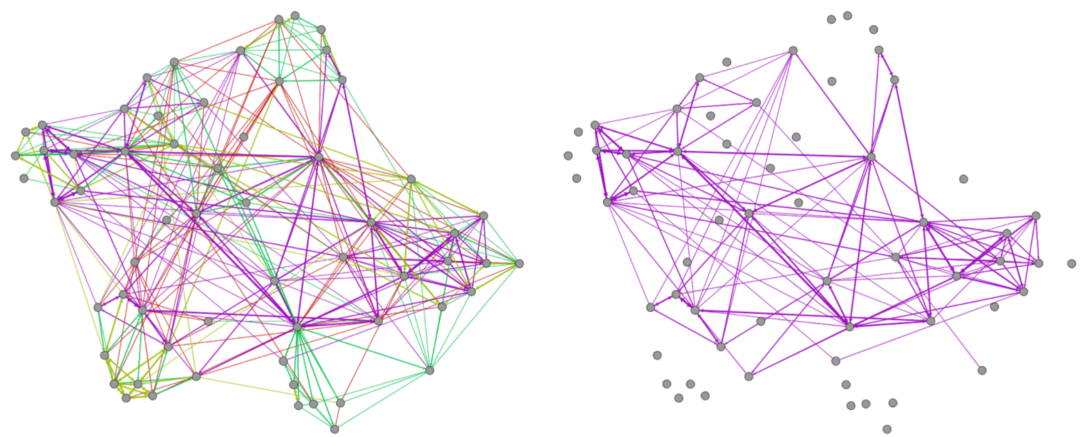
	Work	*Friends*	*Coauthor*	*Lunch*	*Facebook*
# edges	194	88	21	193	124
# con. comp.	2	1	8	1	1
# avg. deg.	6.47	3.74	1.68	6.43	7.75

UPPSALA
UNIVERSITET

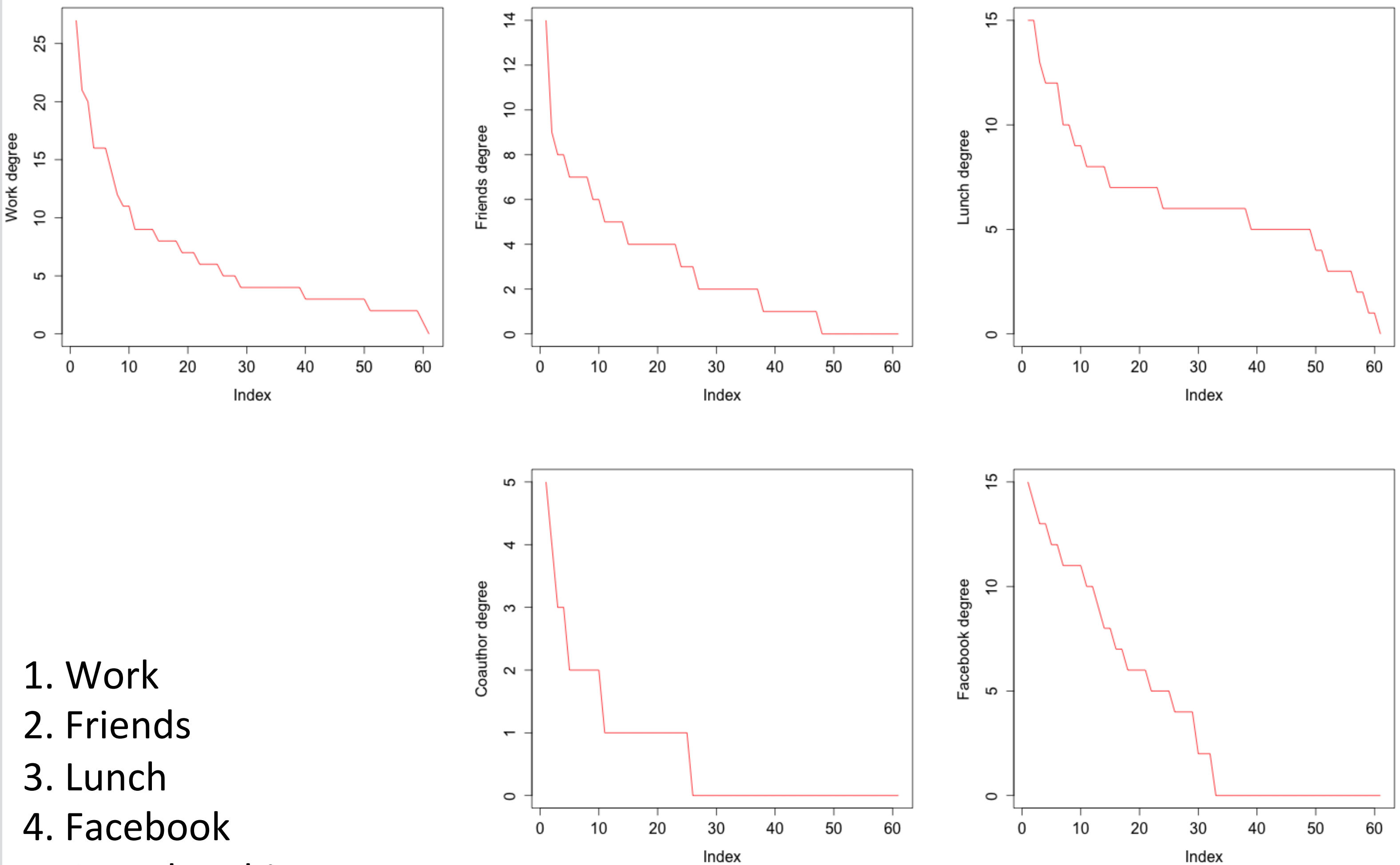
AUCS Dataset



1. Work
2. Friends
3. Lunch
4. Facebook
5. All

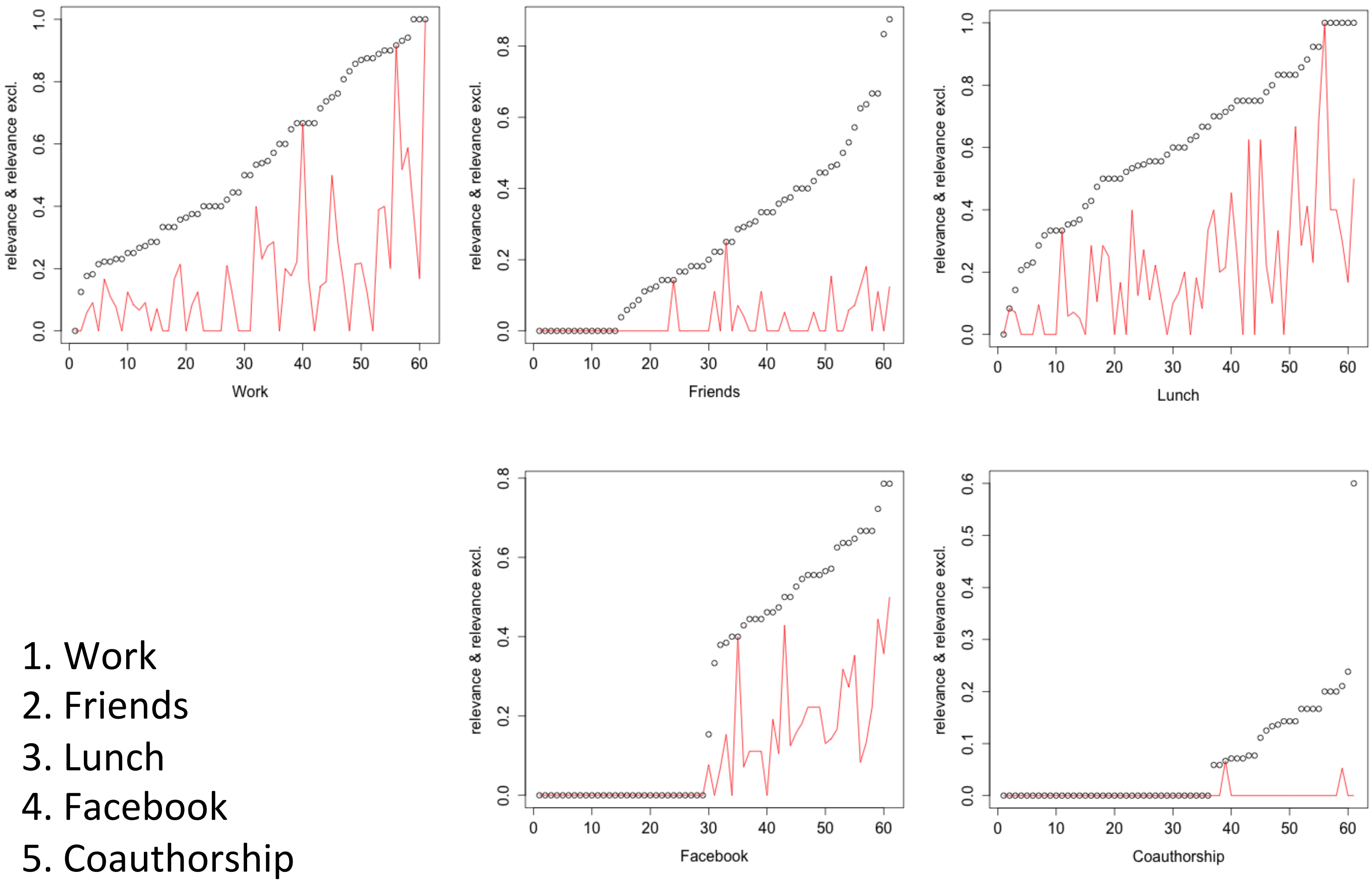


AUCS Dataset



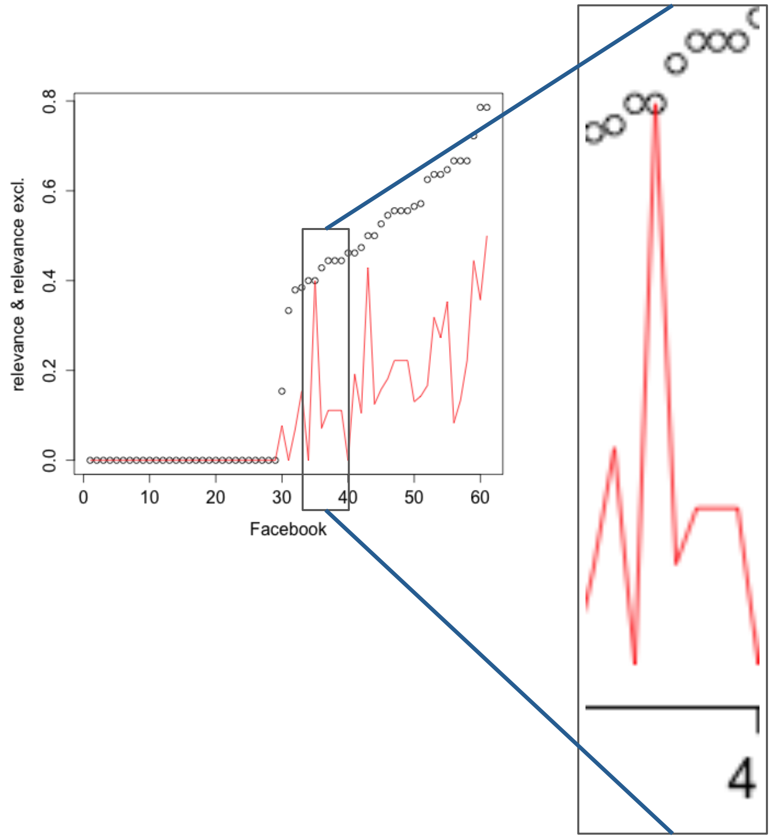
1. Work
2. Friends
3. Lunch
4. Facebook
5. Coauthorship

Relevance and exclusive relevance

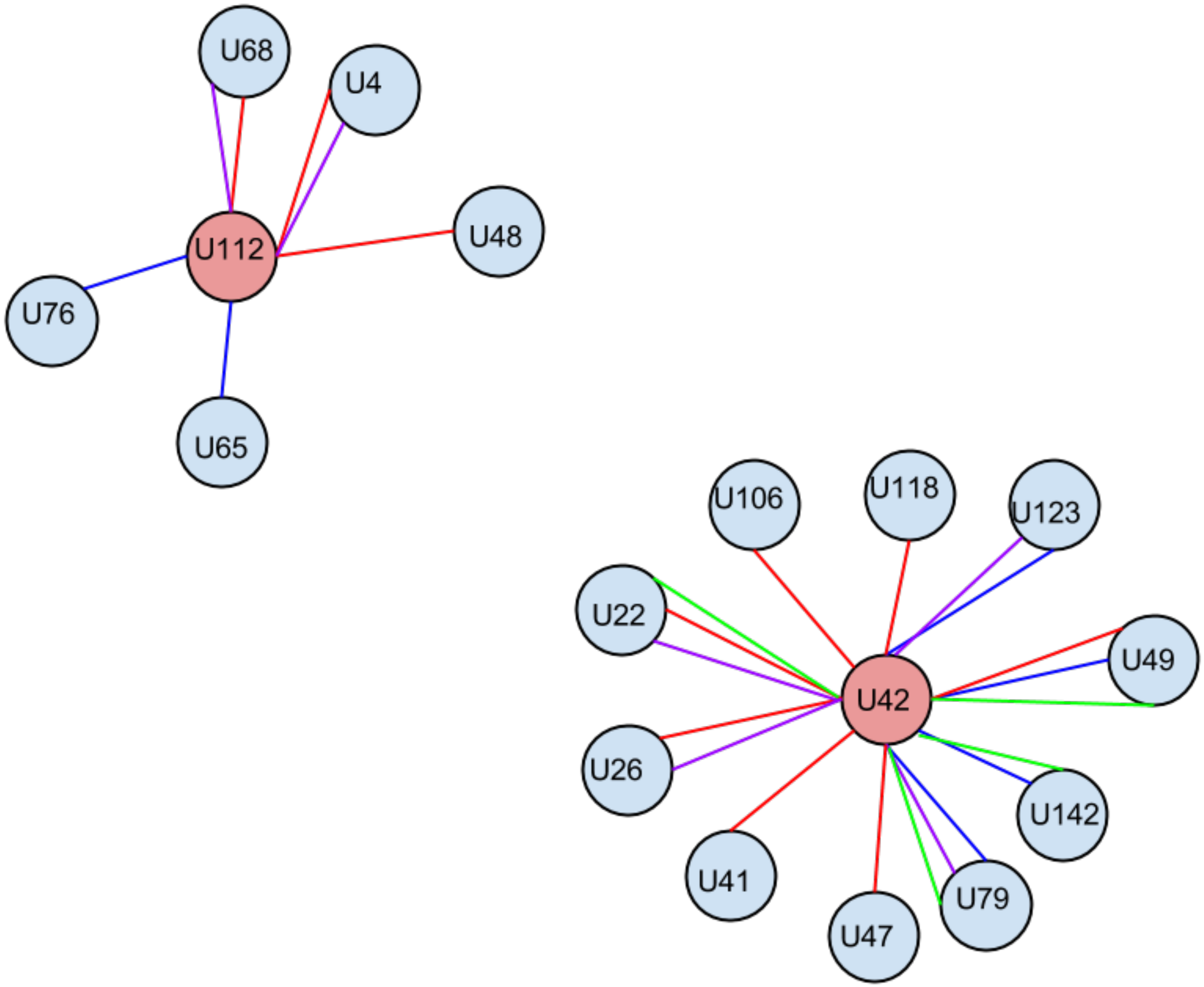


1. Work
2. Friends
3. Lunch
4. Facebook
5. Coauthorship

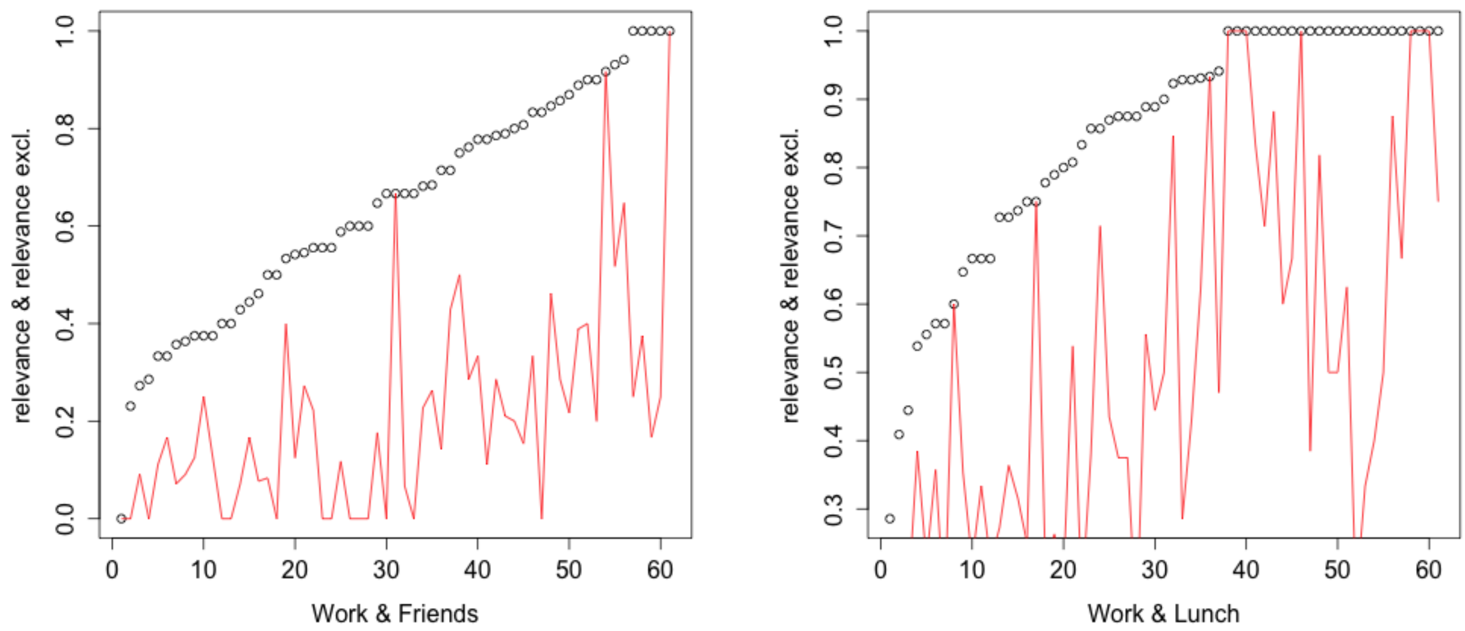
AUCS Dataset



AUCS Dataset



Network complementarity

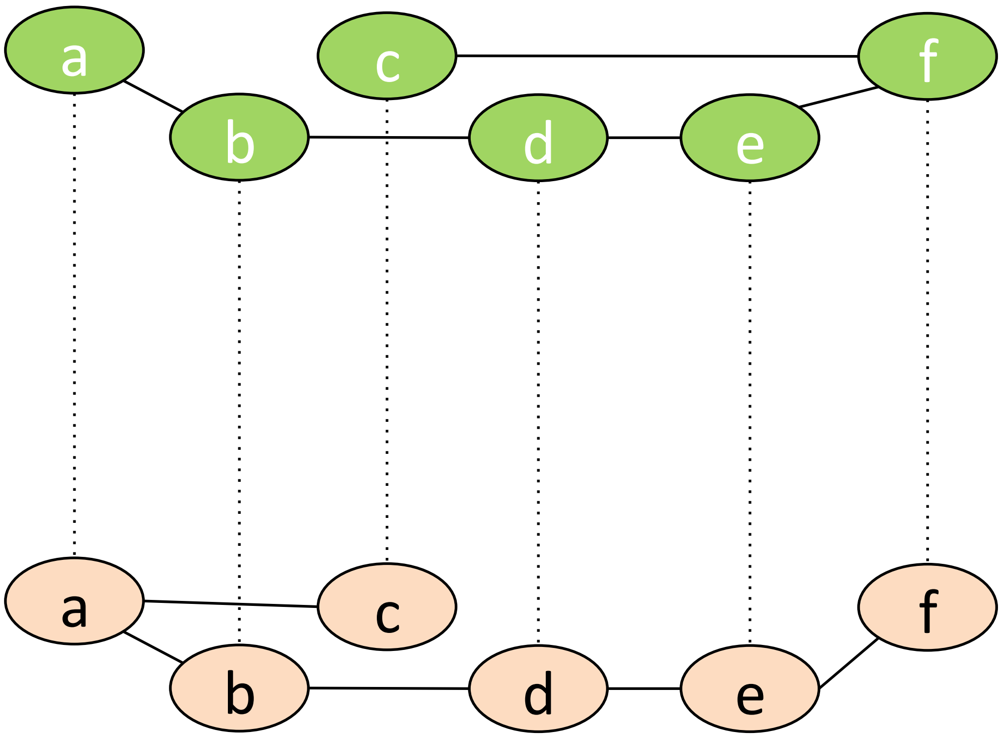


One-minute break

UPPSALA UNIVERSITET

NODE DISTANCE

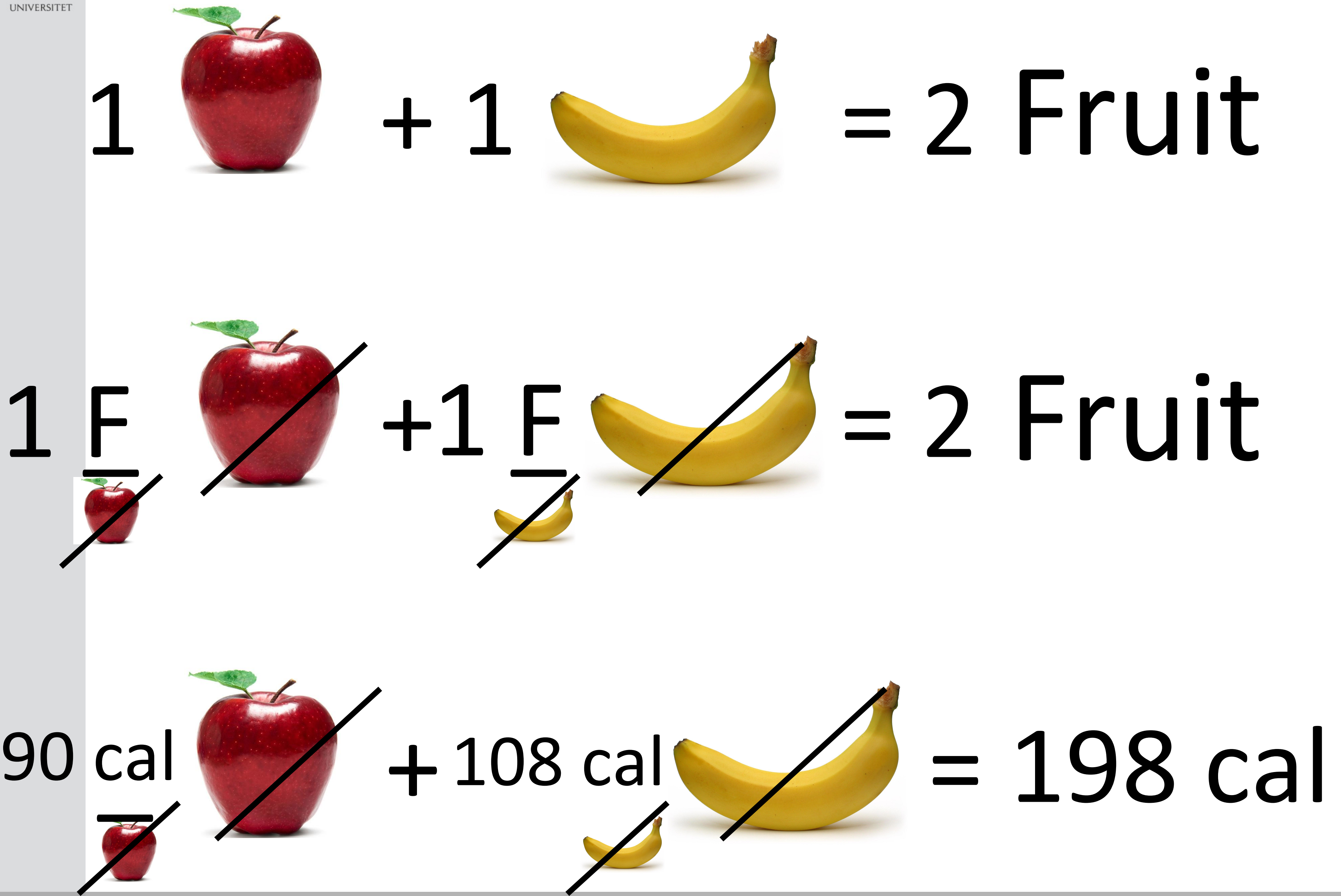
Distances in multi-layer networks



A short digression

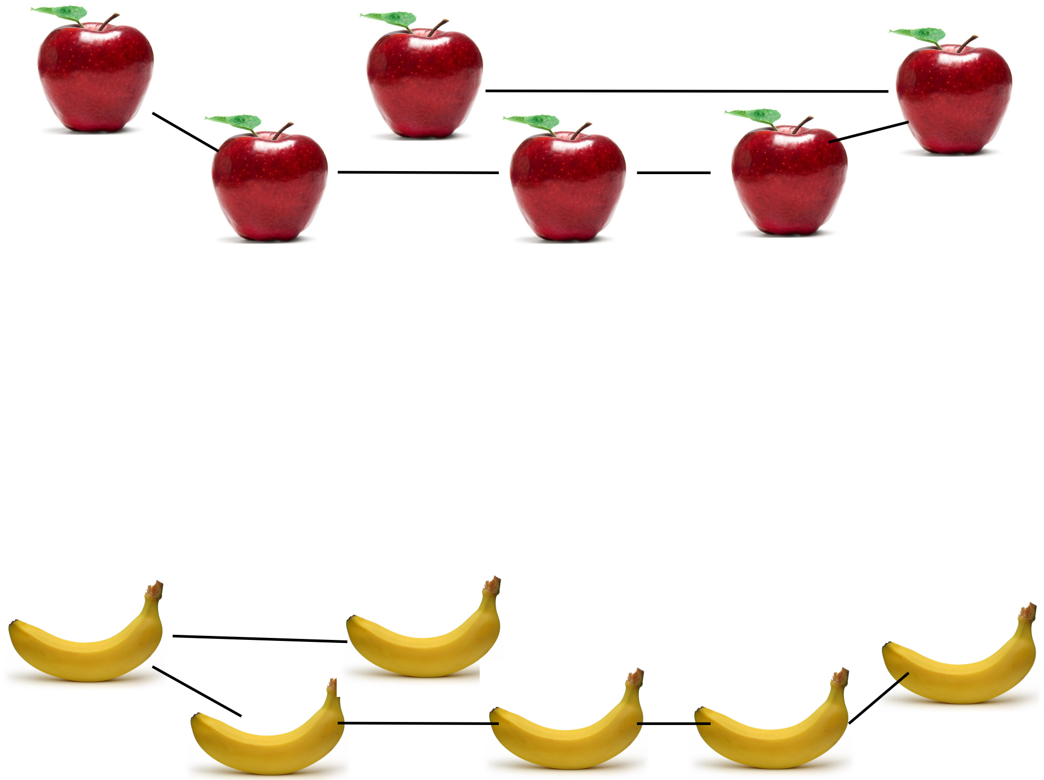
UPPSALA
UNIVERSITET

1  + 1  = 2 Fruit

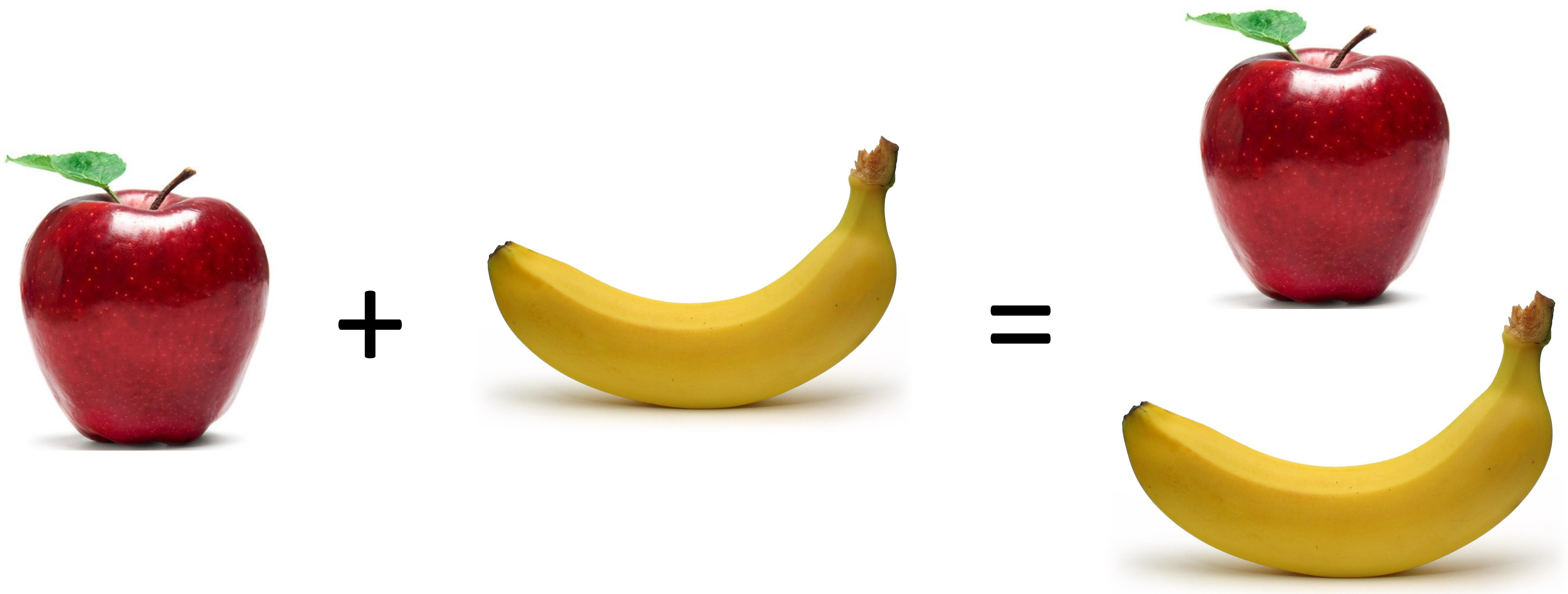

1 + 1 = 2 Fruit
1 F +1 F = 2 Fruit
90 cal + 108 cal = 198 cal

UPPSALA UNIVERSITET

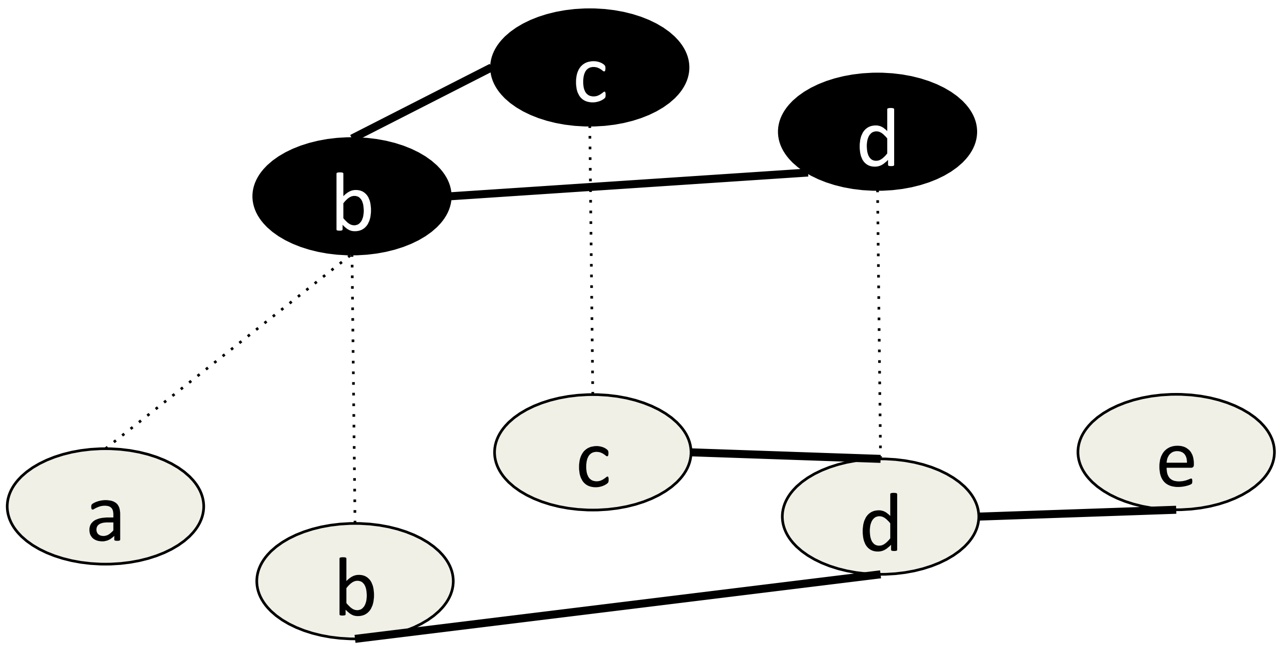
Reduction to single-network distances



Distances in multi-layer networks



Pareto distances

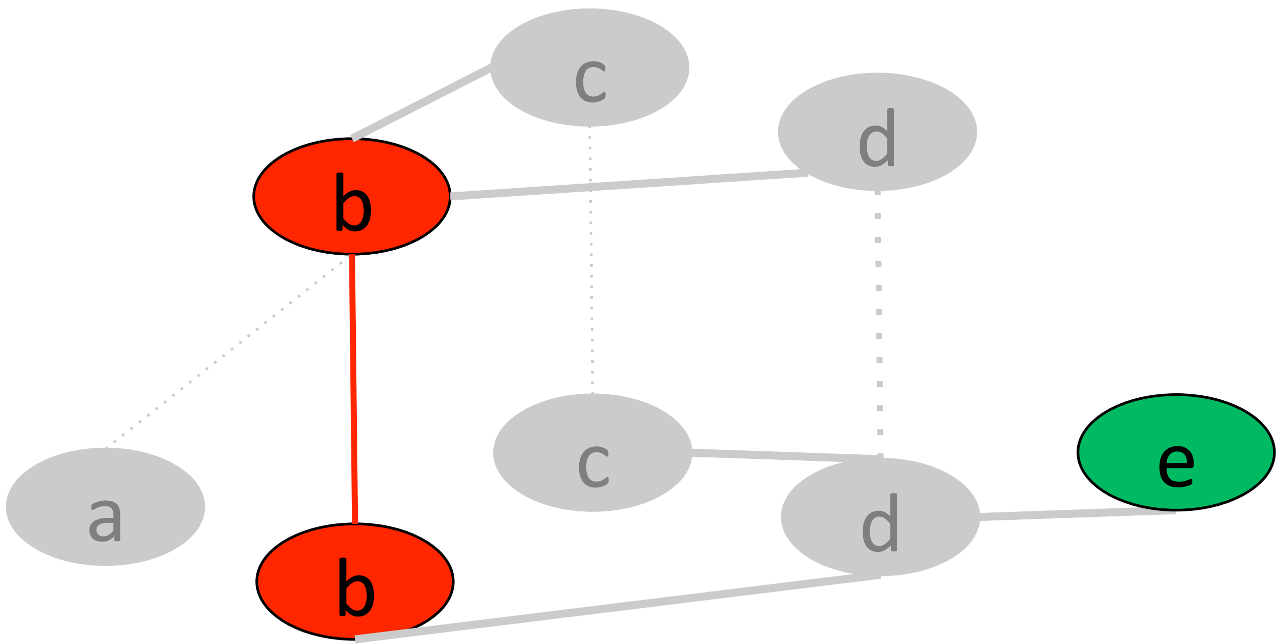


(e.g., LinkedIn)

(e.g., Facebook)

Magnani, M., & Rossi, L. (2013). Pareto Distance for Multi-layer Network Analysis. In SBP (Vol. 7812). Berlin, Heidelberg: Springer

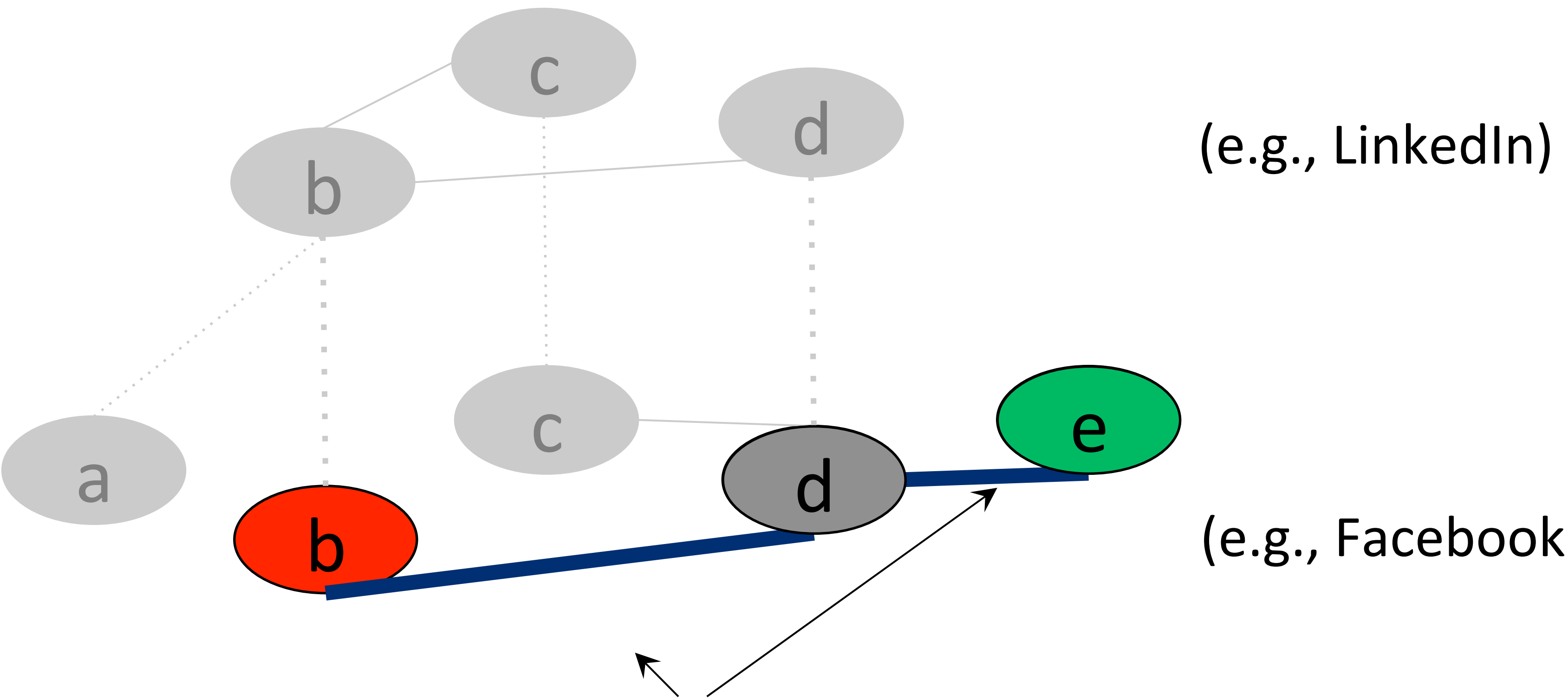
Pareto distances



(e.g., LinkedIn)

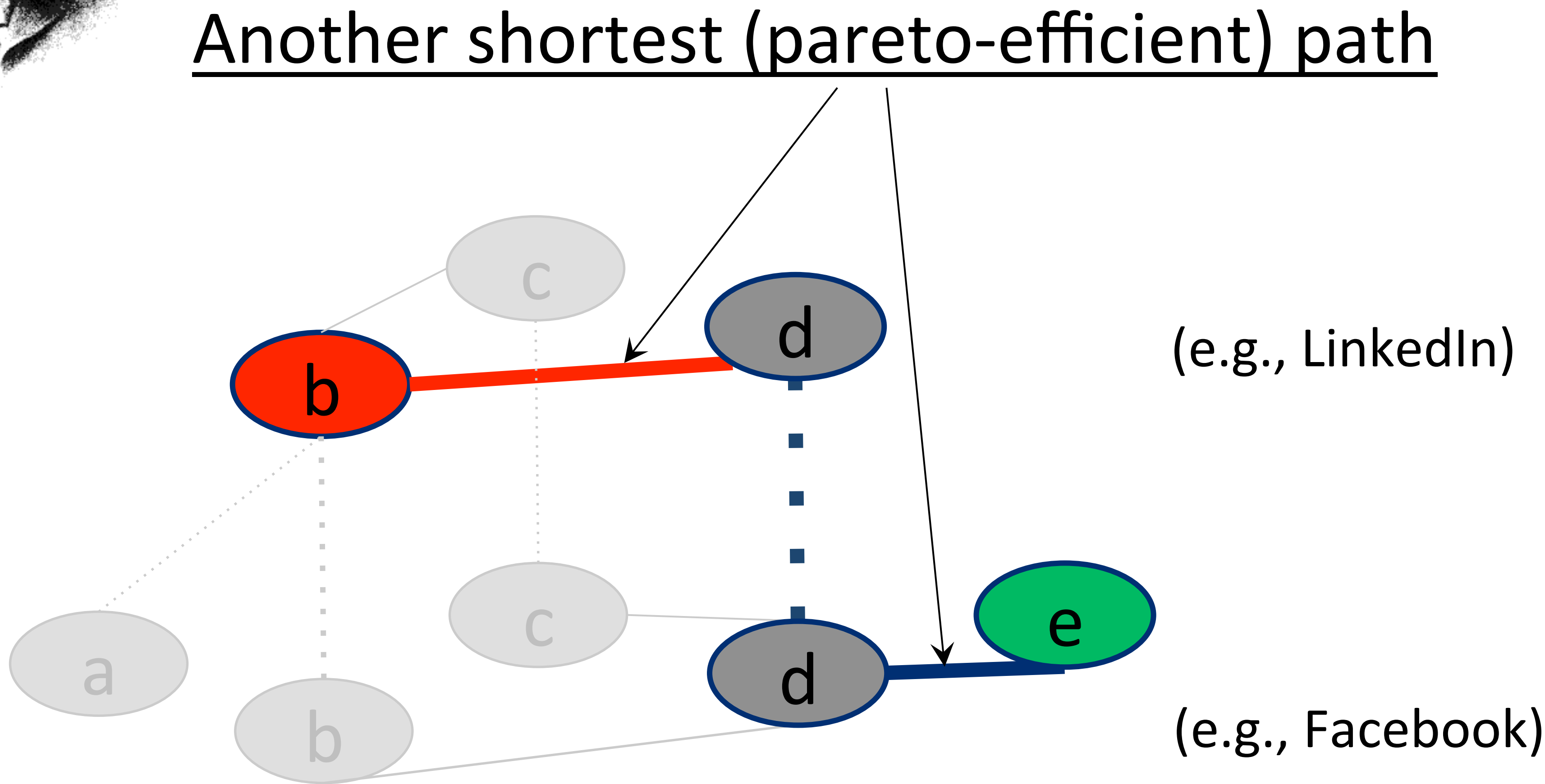
(e.g., Facebook)

Pareto distances



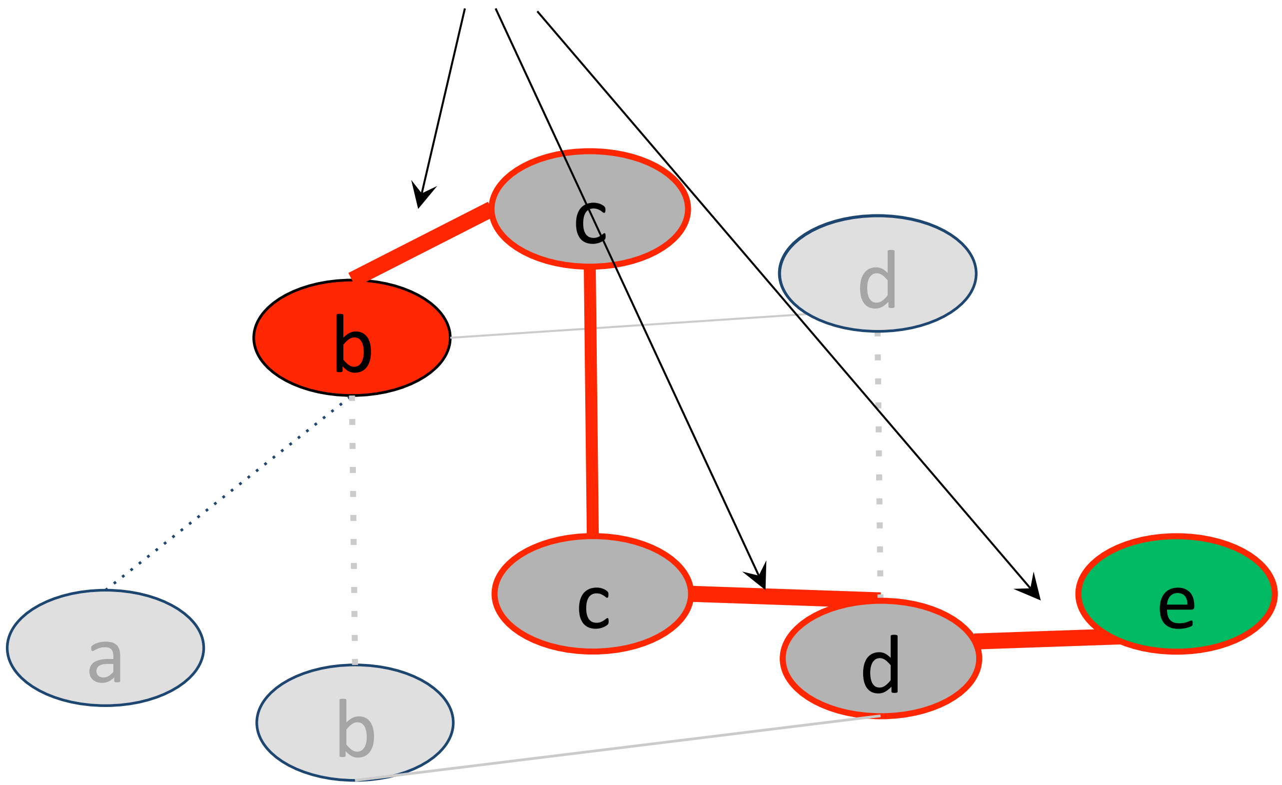
<u>One shortest (pareto-efficient) path</u>

Pareto distances



Pareto distances

One non-pareto-efficient path



(e.g., LinkedIn)

(e.g., Facebook)

Pareto-efficient paths

- Paths not *dominated* by any other path.
- A path dominates another if it is not longer on every single network and is shorter in at least one.

UPPSALA UNIVERSITET

Property 1

Every shortest path in any possible flattened network is a pareto-optimal multi-layer path.

Property 2

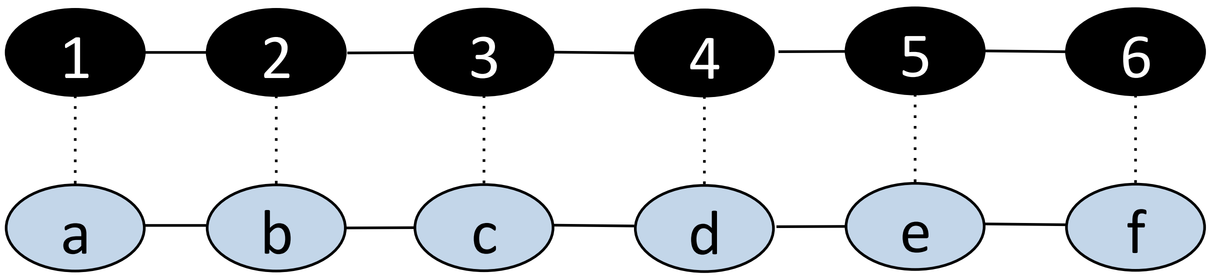
Every pareto-optimal multi-layer path is a shortest path in at least one flattened network.

UPPSALA UNIVERSITET

Property 3

On a single network,
the two concepts correspond.

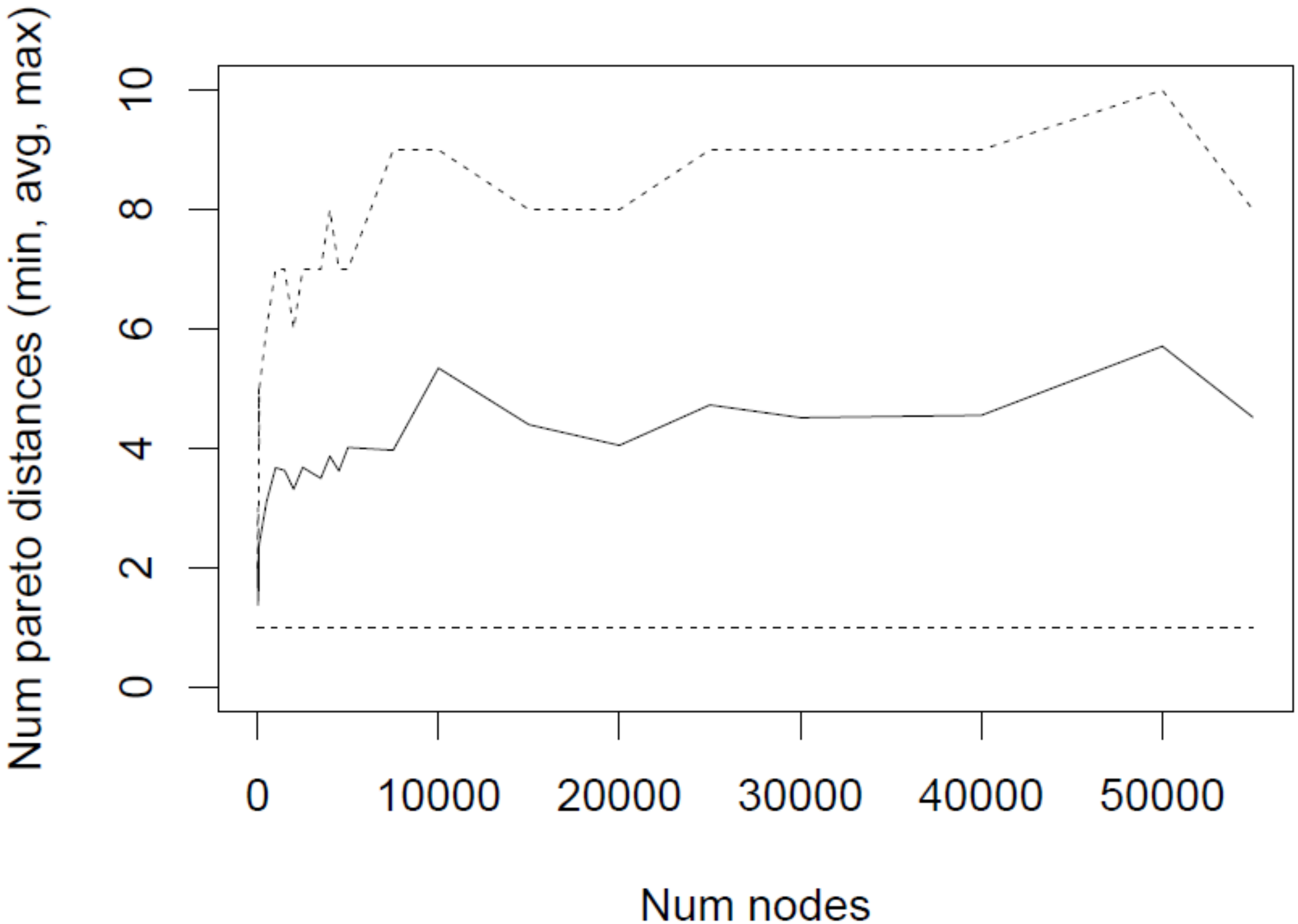
Complexity: worst-case scenarios



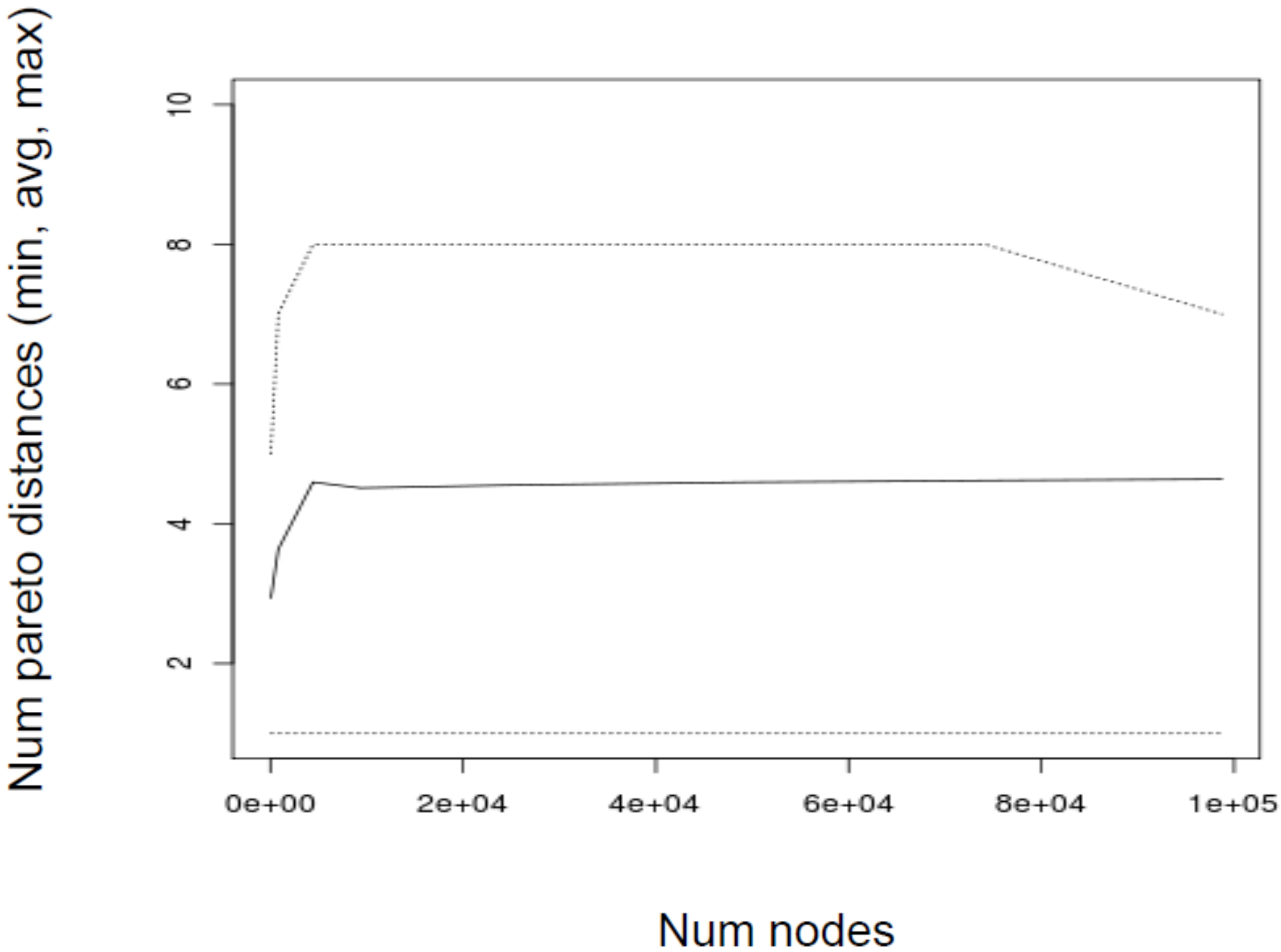
Scalability analysis

- Test the growth of Pareto distances between any two nodes with increasing network size.
- Synthetic data.
  - Up to 50 000 nodes per network.
  - About 150 000 undirected edges.
- Real data.
  - About 150 000 users on Twitter and Friendfeed.
  - About 20 000 000 directed edges.

Cardinality vs. Size (synthetic data)

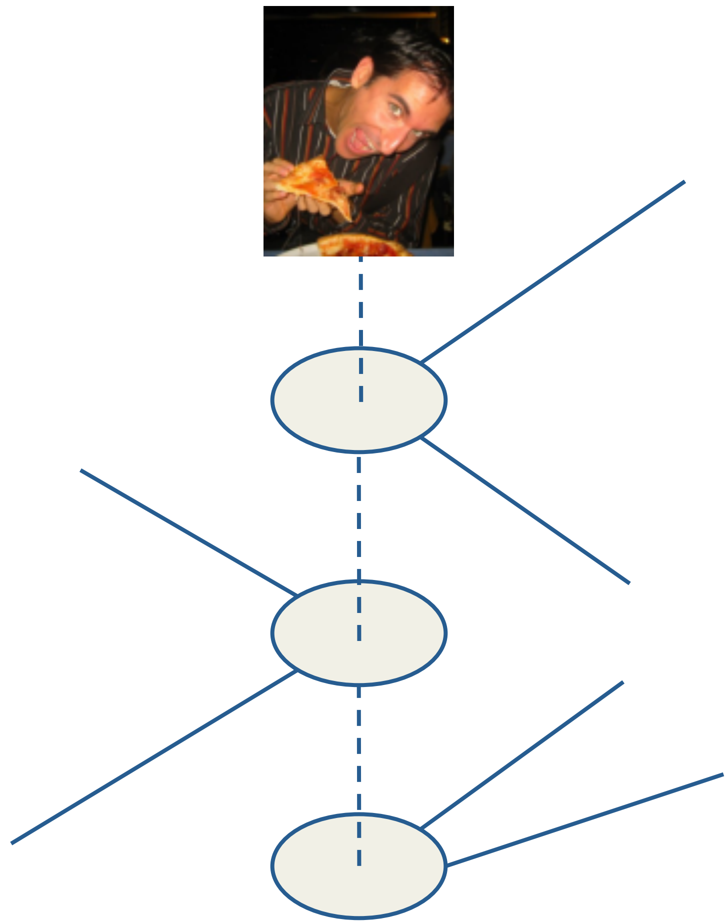


Cardinality vs. Size (real data)



DISTANCES: EXAMPLES

Applying Pareto distance to AUCS



Distance to... U109

Interpretation

U54 **-lunch-** U109

U54 **-facebook-** U109

U54 **-friend-** U109

U54 **-work-** U109

close friend
connected through
several networks

Distance to... U109

U54 **-lunch-** U3

U54 **-facebook-** U3

U54 **-friend-** U3

U54 **-work-** U90 **-work-** U3

Interpretation

Close friend but not a coworker

UPPSALA UNIVERSITET

Distance to... U19

U54 **-work-** U19

U54 **-facebook-** U10 **-lunch-** U19

U54 **-friend-** U10 **-lunch-** U19

U54 **-facebook-** U79 **-friend-** U73 **-friend-** U19

U54 **-friend-** U79 **-friend-** U73 **-friend-** U19

U54 **-lunch-** U76 **-coauthor-** U130 **-lunch-** U32 **-lunch-** U73 **-lunch-** U19

U54 **-lunch-** U79 **-lunch-** U130 **-lunch-** U32 **-lunch-** U73 **-lunch-** U19

Interpretation: Not all coworkers are friends

Wrapping up & Open Problems

- Centrality:
  - Extended measures considering the multiple interdependent networks.
  - New network-specific measures.
- Computing social distances.
- Possible applications:
  - Multi-layer betweenness.
  - Community detection.
- Other measures?
- Methodology?

UPPSALA UNIVERSITET

FORMATION OF MULTIPLE NETWORKS

Evolution Model → Synthetic data

Evolution Model

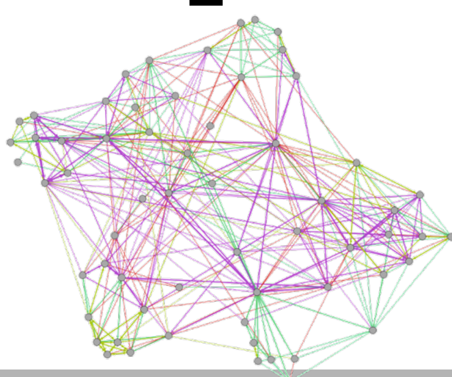
MODEL PARAMETERS

→ Properties

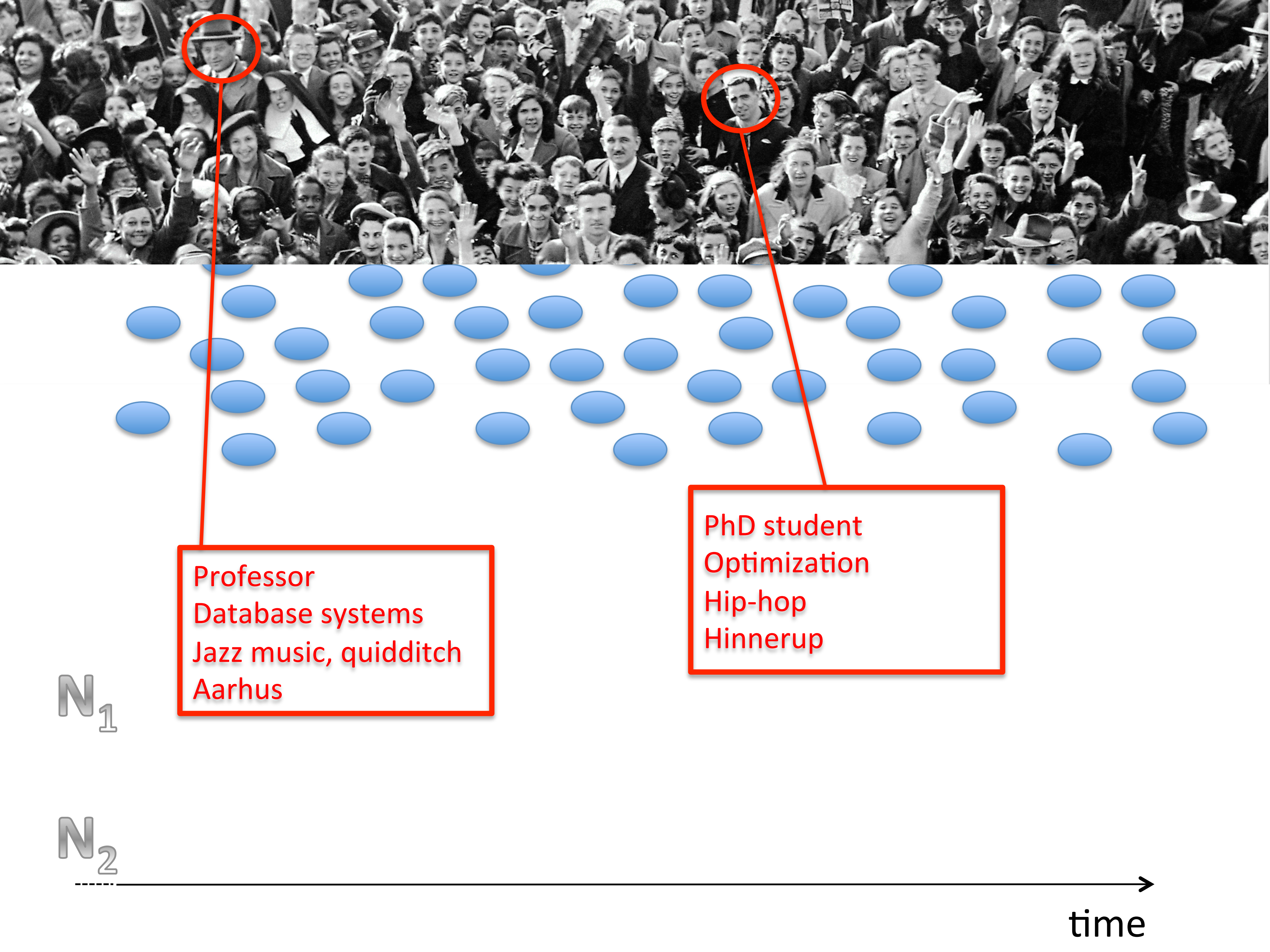
Evolution Models

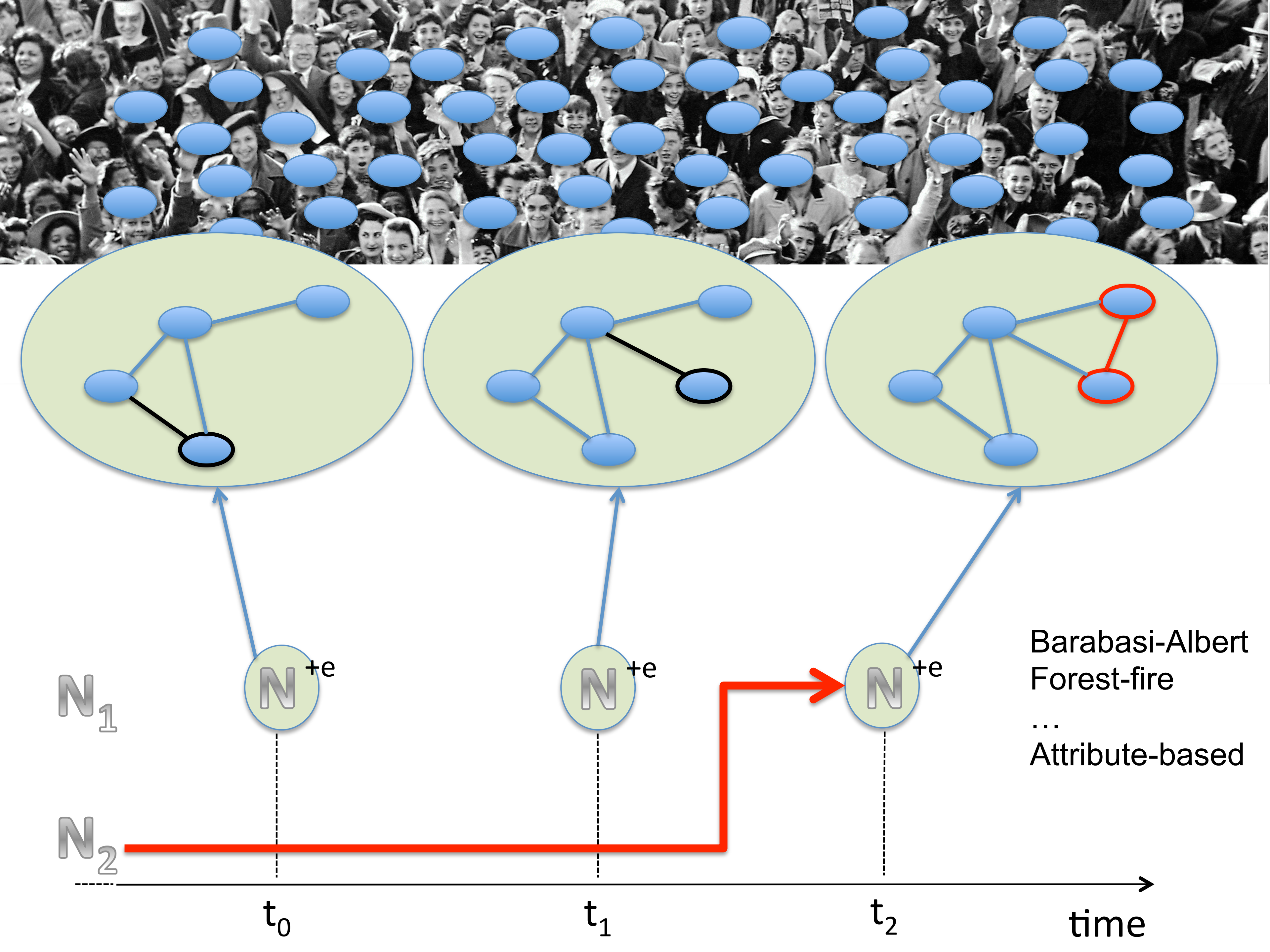
PARAMETERS (INCLUDING MODELS)

→ FIT → Explanation

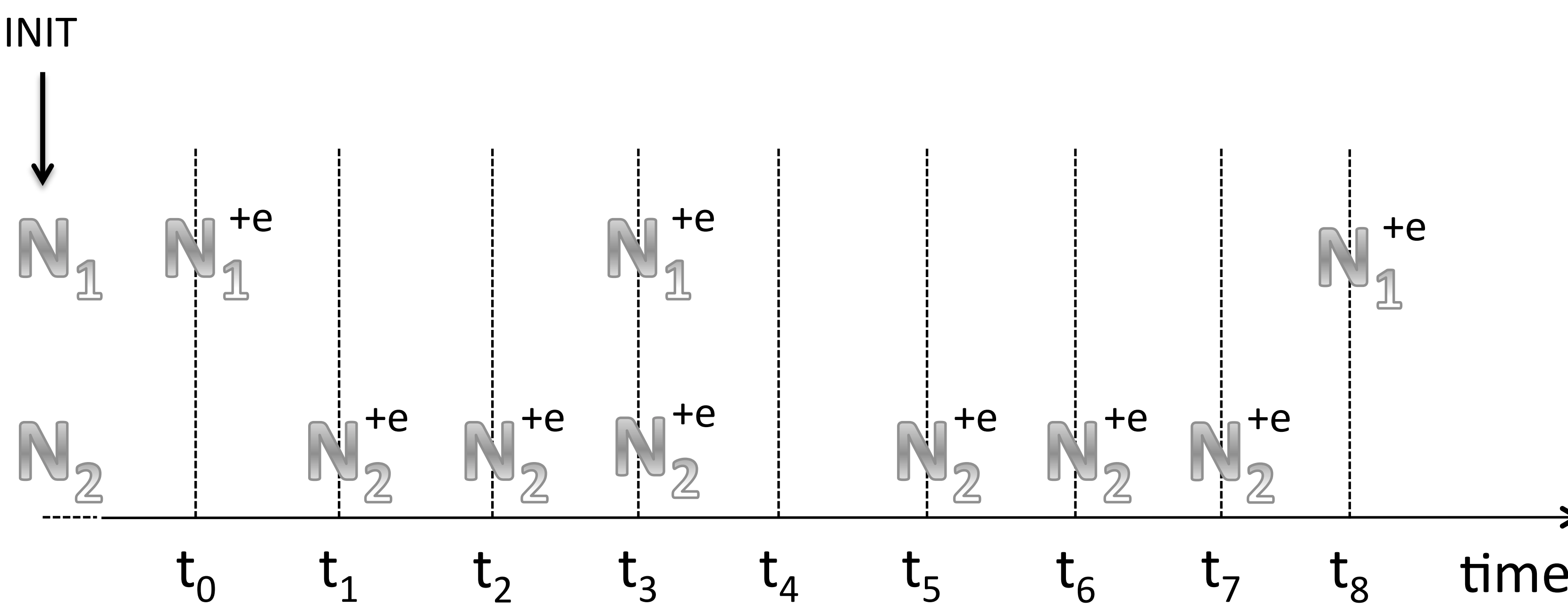


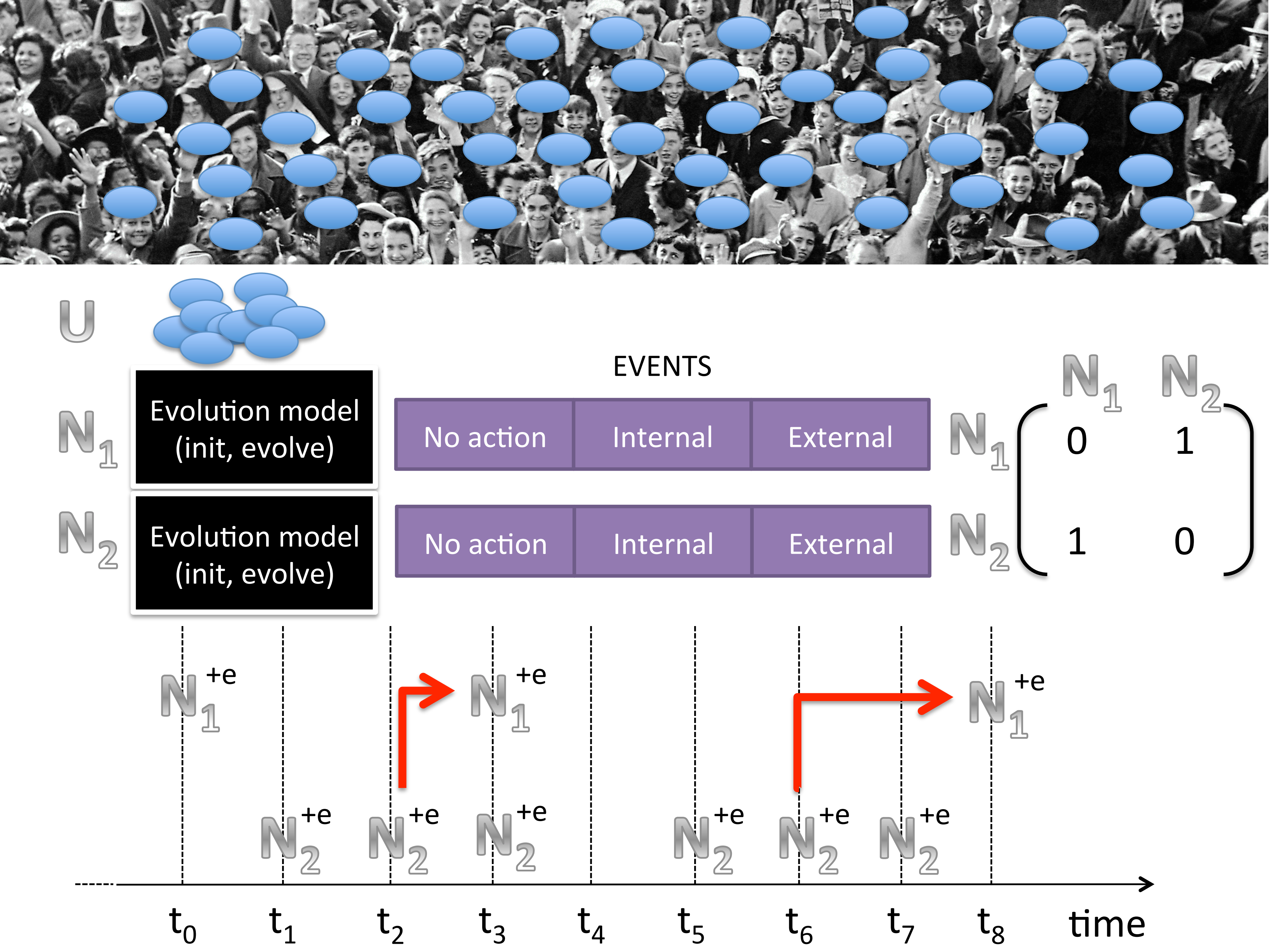
Study framework


Professor
Database systems
Jazz music, quidditch
Aarhus
PhD student
Optimization
Hip-hop
Hinnerup
N_1
N_2
time



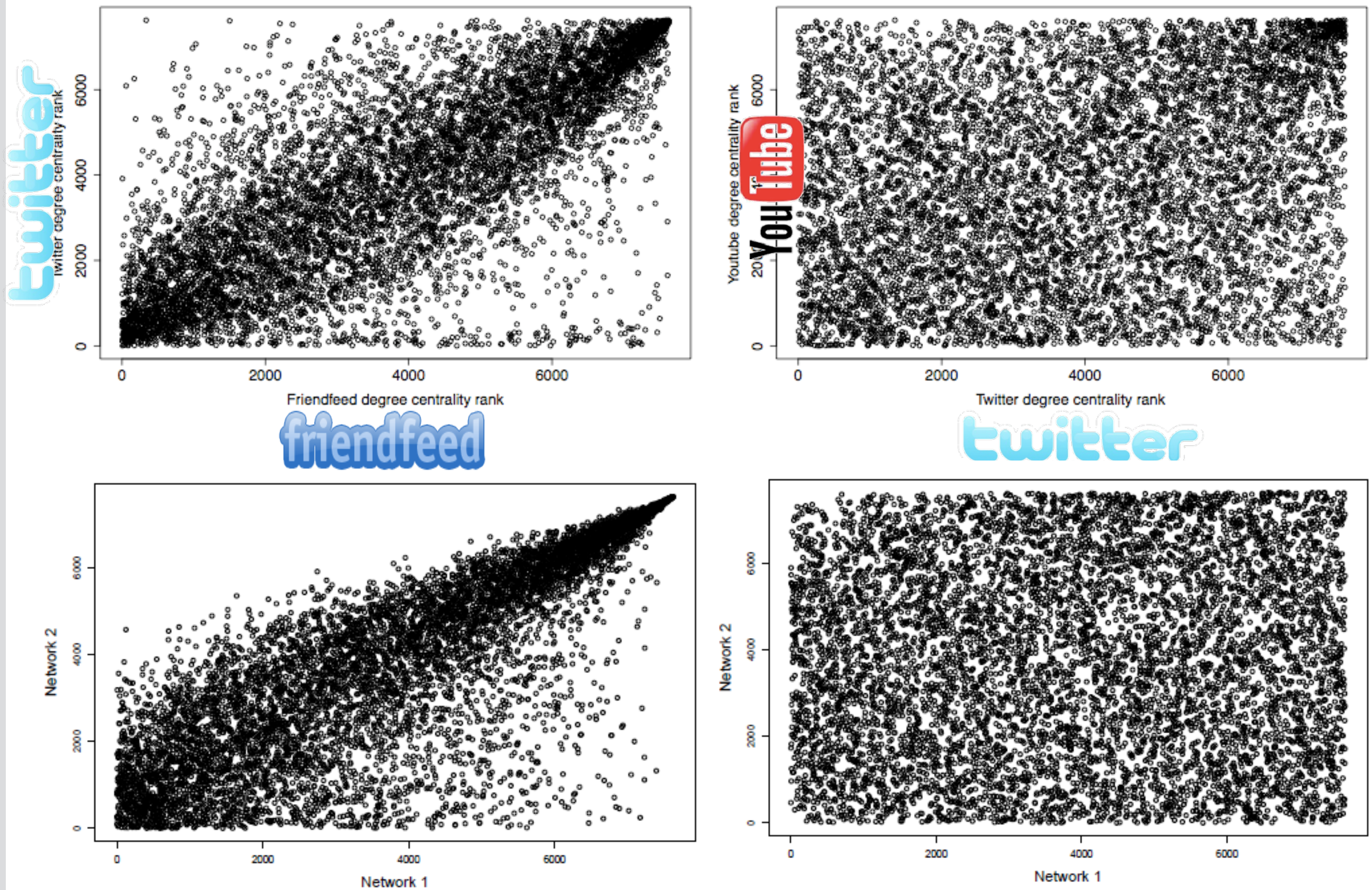
+e
+e
+e
N_1
N_2
Barabasi-Albert
Forest-fire
…
Attribute-based
t_0
t_1
t_2
time


INIT
N1
N2
N1 +e
N2 +e
N2 +e
N1 +e
N2 +e
N2 +e
N2 +e
N2 +e
N1 +e
t0
t1
t2
t3
t4
t5
t6
t7
t8
time


U
N1
Evolution model
(init, evolve)
N2
Evolution model
(init, evolve)
EVENTS
No action
Internal
External
No action
Internal
External
N1
N2
N1
N2
0
1
1
0
N1 +e
N2 +e
N2 +e
N1 +e
N2 +e
N2 +e
N2 +e
N2 +e
N1 +e
t0
t1
t2
t3
t4
t5
t6
t7
t8
time

Data

Friendfeed 37.997 edges
Twitter 67.123 edges
YouTube 1.185 edges



Compairing Degree centrality ranks on Multiple network Real Data

Wrapping up & Open problems

- Identified a minimal set of parameters to control basic co-evolution patterns.
  - Synchronization / different growth.
  - Internal/external dynamics.
- Future research questions.
  - Relevant configurations for > 2 networks?
  - How different formation models for different networks interact?

CLUSTERING (COMMUNITY DETECTION)

Community detection

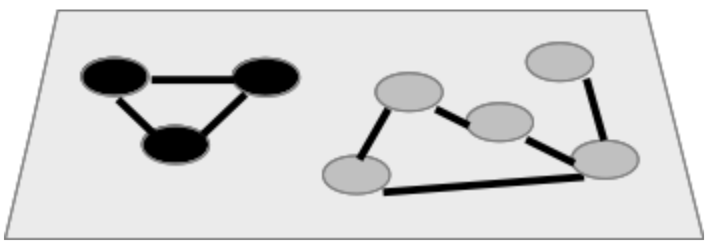
- A large amount of research for single networks (see S. Fortunato's survey).
- However, more questions than answers.
- A relevant one: what is a community?

A few basic definitions

- Community based on network topology.
- How to quantify the connectedness?
- We need a quality function to optimize.
- Two relevant concepts:
  - Modularity.
  - (Quasi-)clique.
- Introduce modularity for single graphs, then extend it to multiple graphs.
- Give an overview over clique-based methods and draw a future scenario.

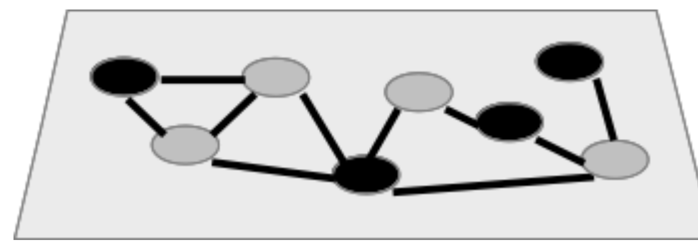
Modularity

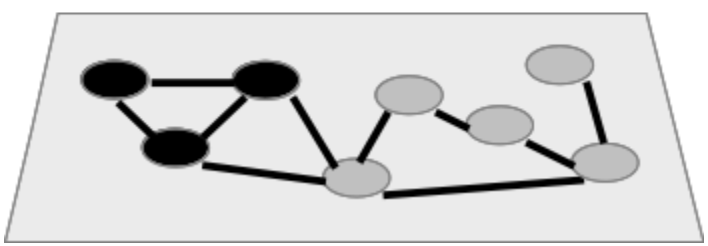
$$Q = \frac{1}{2m}\sum_{ij}\left(a_{ij} - \frac{k_i k_j}{2m}\right)\delta(\gamma_i, \gamma_j)$$



Q = 0.47

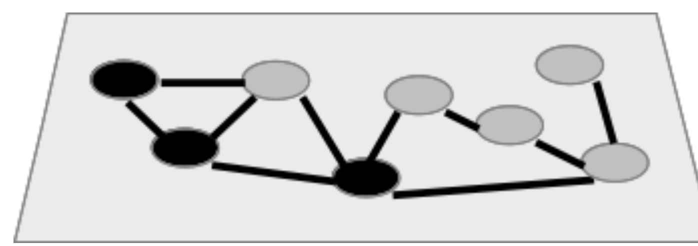
Q = -0.4





Q = 0.28

Q = 0



UPPSALA
UNIVERSITET

Modularity

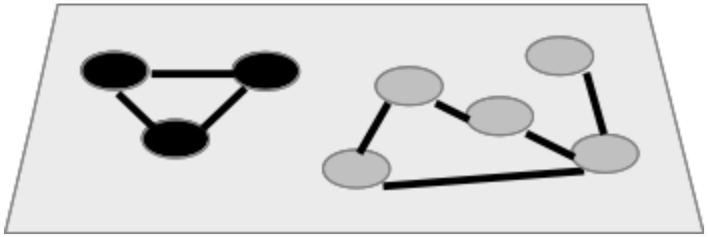
Divided by number of edges

Edge between nodes

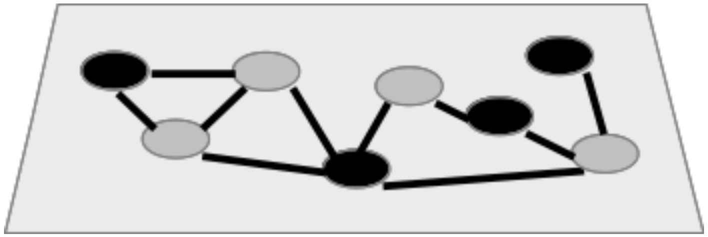
null model

Only consider if nodes inside same community

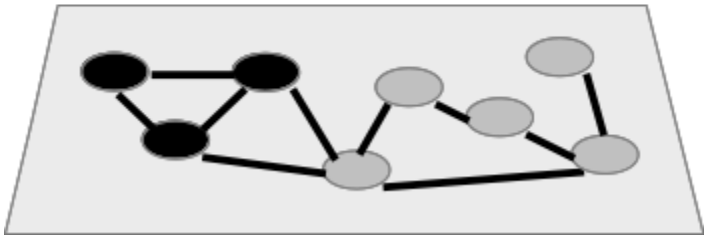
$$Q = \frac{1}{2m}\sum_{ij}\left(a_{ij} - \frac{k_i k_j}{2m}\right)\delta(\gamma_i, \gamma_j)$$



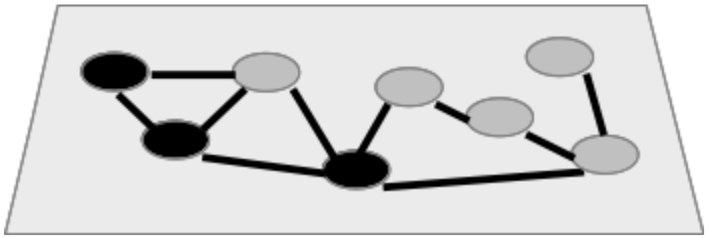
Q = 0.47



Q = -0.4

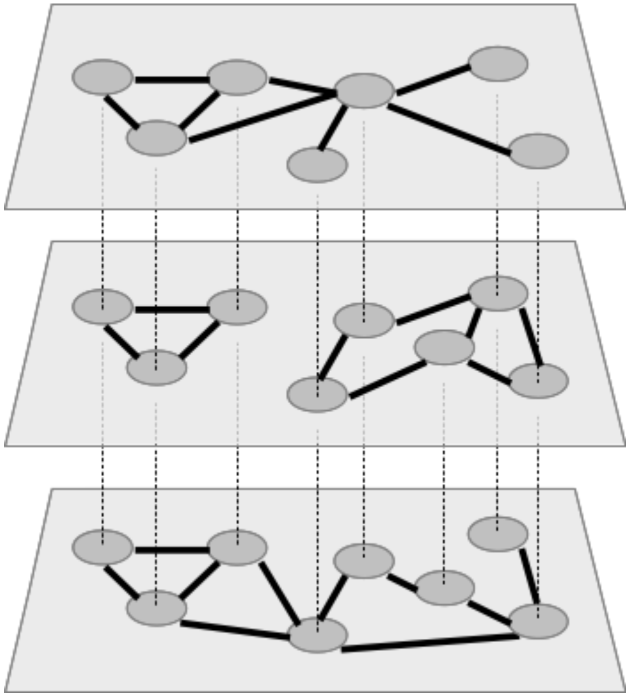


Q = 0.28



Q = 0

Multiplex modularity



$$Q = \frac{1}{2m}\sum_{ij}\left(a_{ij} - \frac{k_i k_j}{2m}\right)\delta(\gamma_i, \gamma_j)$$

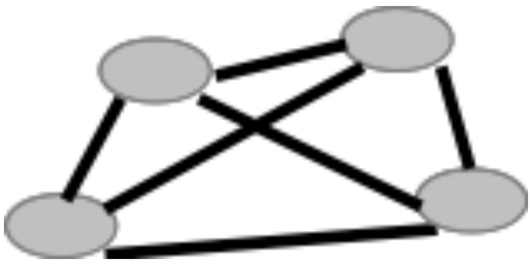
$$Q_m = \frac{1}{2\mu}\sum_{ijsr}\left[\left(a_{ijs} - \frac{k_{is}k_{js}}{2m_s}\right)\delta(s,r) + c_{jsr}\delta(i,j)\right]\delta(\gamma_{i,s}, \gamma_{j,r})$$

nodes on same network

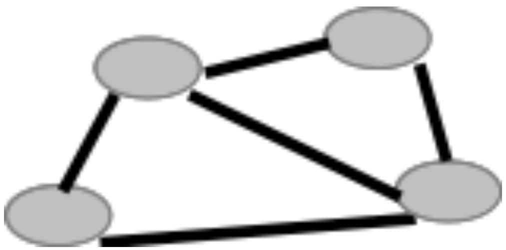
same node on different networks

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. Science (New York, N.Y.),328(5980).

Clique finders

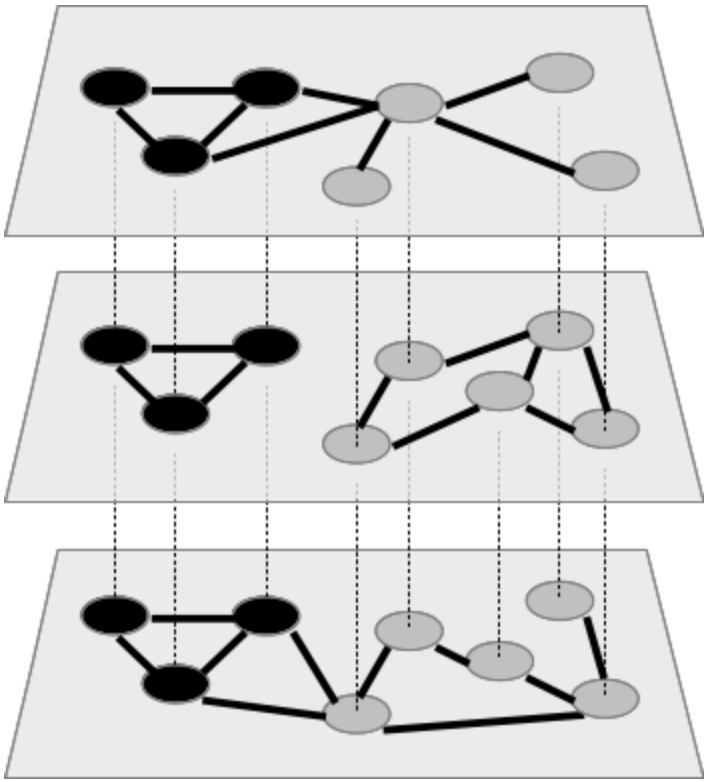


Clique



Quasi-Clique
(>50%)

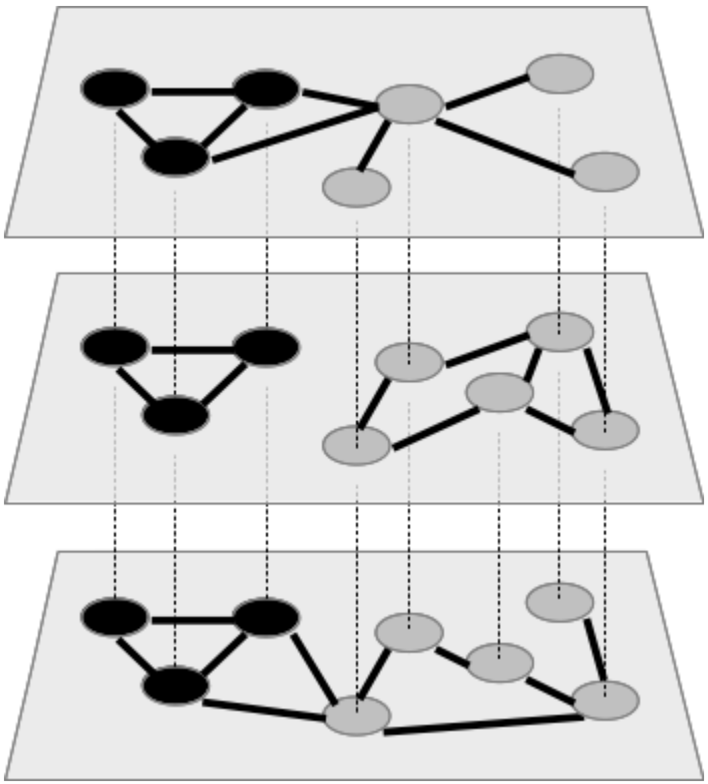
Clique finders (top-down)



Present in at least s% of the networks (support)

Zhiping Zeng, J. W. (2006). Coherent closed quasi-clique discovery from large dense graph databases. 12th International Conference on Knowledge and Data Discovery (KDD), 2006

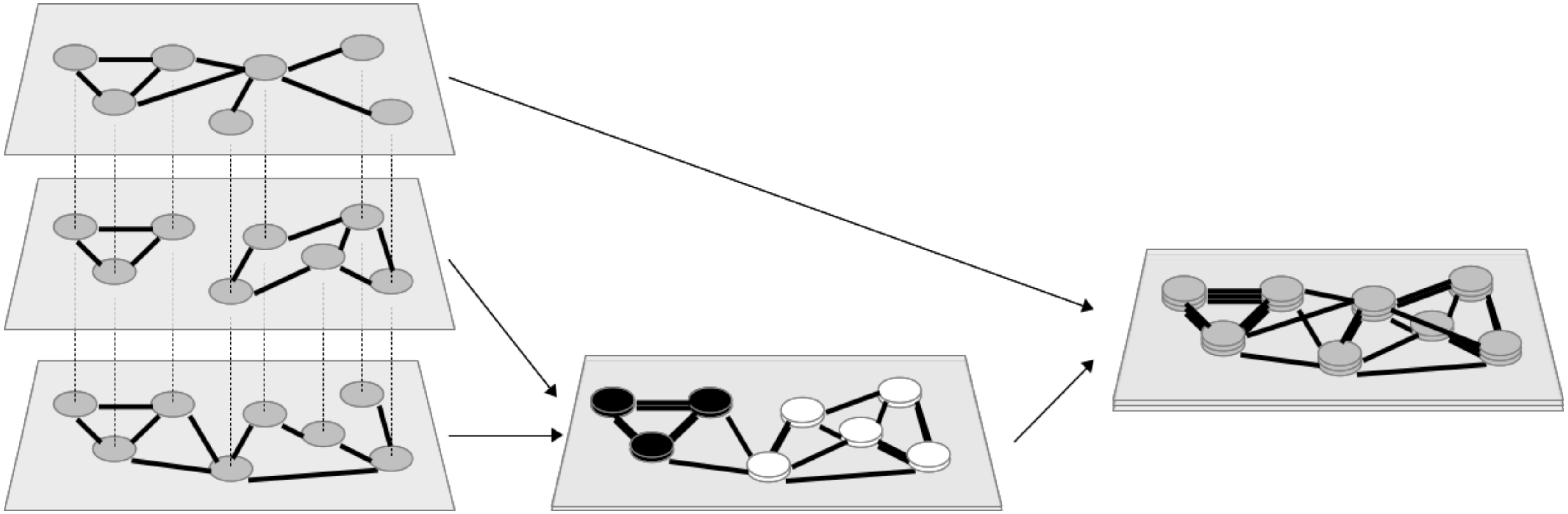
Clique finders: bottom-up



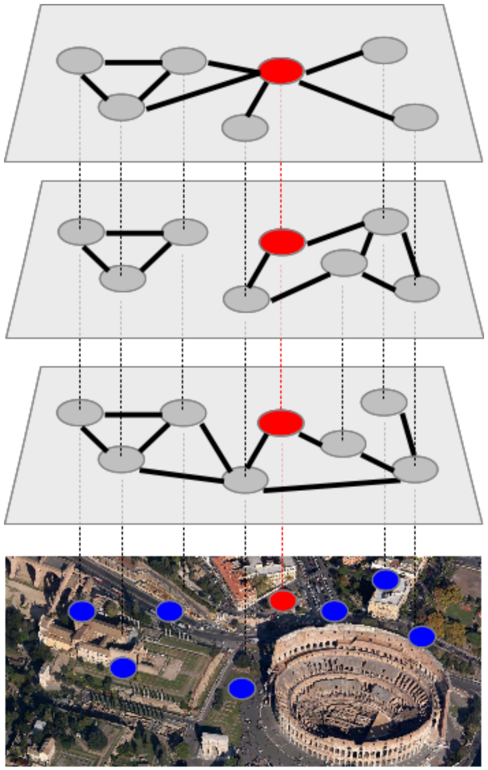
Present in any number of graphs
(but in all of them)

Boden, B., Günnemann, S., Hoffmann, H., & Seidl, T. (2012). Mining coherent subgraphs in multi-layer graphs with edge labels. Proceedings of the 18th ACM SIGKDD international conference.

What's next: emerging clusters?



SOME PRACTICAL REMARKS



friendfeed

YouTube

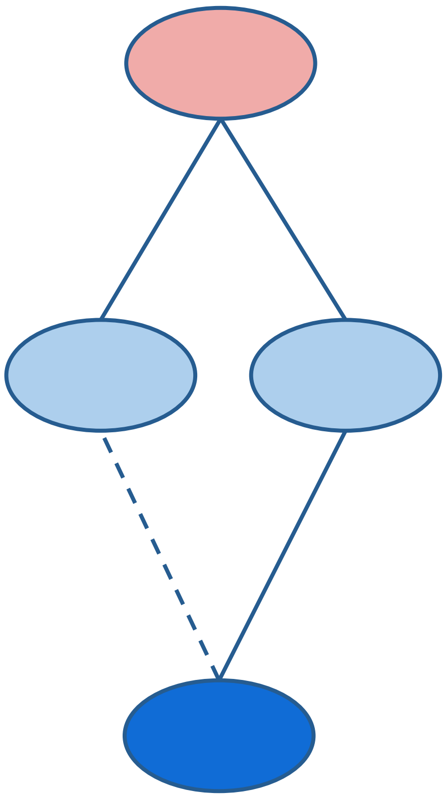
twitter

off-line
encounters

data collection
identity mapping
privacy issues

heterogeneous data (and the lack of proper multiplex archives) lead toward mixed data collection methods: *scalability & reliability* problems

UPPSALA
UNIVERSITET



(LinkedIn)

(Twitter)

(Facebook)

multiplex identity can show unexpected and evolving structure.

UPPSALA
UNIVERSITET

A large number of layers gives a high level of additional information about users making harder real and effective anonymization (e.g. T3 dataset)

Related research projects

http://sigsna.net/impact

http://www.multiplexproject.eu/

http://www.plexmath.eu/

http://lasagne-project.eu

Related topics not covered here

- Mining heterogeneous information networks
  - J. Han's group work
- Link prediction
  - E.g., Rossetti et al., Scalable link prediction on multidimensional networks. ICDMW, 2011.
- Multiple network visualization
  - Very little work.
  - Dai et al., ViStruclizer: A Structural Visualizer for Multi-dimensional Social Networks. PAKDD, 2013