

Users are a Complex Phenomenon

Julia Neidhardt

Research Unit of E-Commerce

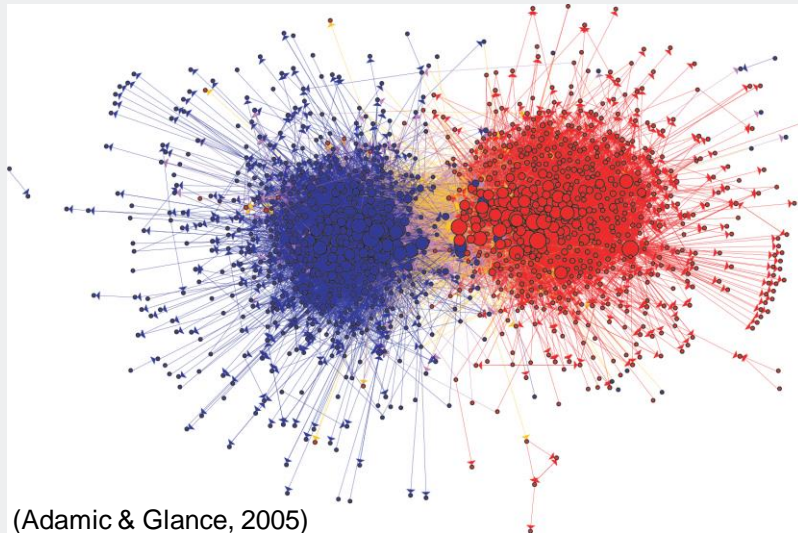
TU Wien, Austria

julia.neidhardt@tuwien.ac.at

<http://www.ec.tuwien.ac.at/neidhardt/>

Background

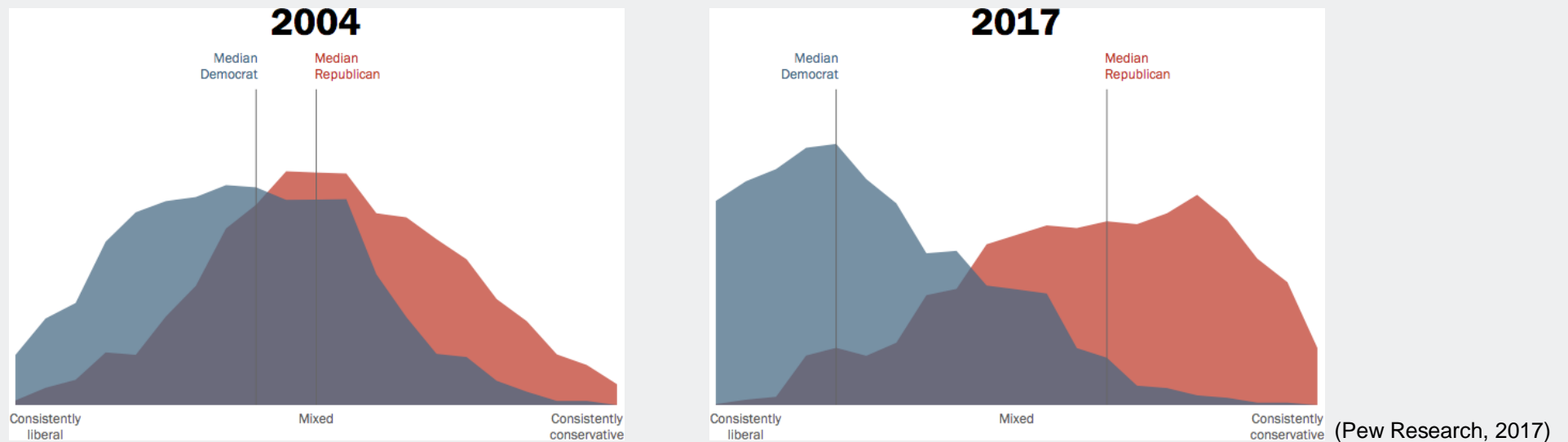
- In social networks, individuals are often connected to others who are similar with respect to socio-demographics, political values, etc. (i.e., homophily)
- Big picture: the overall social structure is typically comprised of separated, stable groups with hardly any overlap



One famous example: network structure of political bloggers (Democrats vs. Republicans) during US presidential elections in 2004

Background (2)

- At the same time, people often have mixed opinions on specific topics (e.g., immigrants & gay rights)
 - This level is highly dynamic;
 - But: positions are getting more and more polarized.



→ Studies suggest that social media reinforce polarization, not only in the US but globally (Kozinets et al., 2016)

Background (3)

- Online dynamics and questions regarding social influence have been gaining increasing attention in public discourse
 - Buzzwords: fake news, micro targeting, filter bubbles, echo chambers, hate speech, etc.
 - Phenomena like this challenge the vision of the web as an open platform
(Berners-Lee, 2017)
- Thus, studying these online dynamics as well as social influence processes might be crucial to understand why people adopt certain opinions/beliefs
- However, a comprehensive, multi-level user model is required

Modelling Social Systems

- Common approaches to model social systems:
 - 1. Individual level**
 - Independent individuals
 - 2. Group level**
 - Sets of individuals
 - 3. Network level**
 - Dyadic relations
- **Now:** discuss these levels and why it is important to integrate them

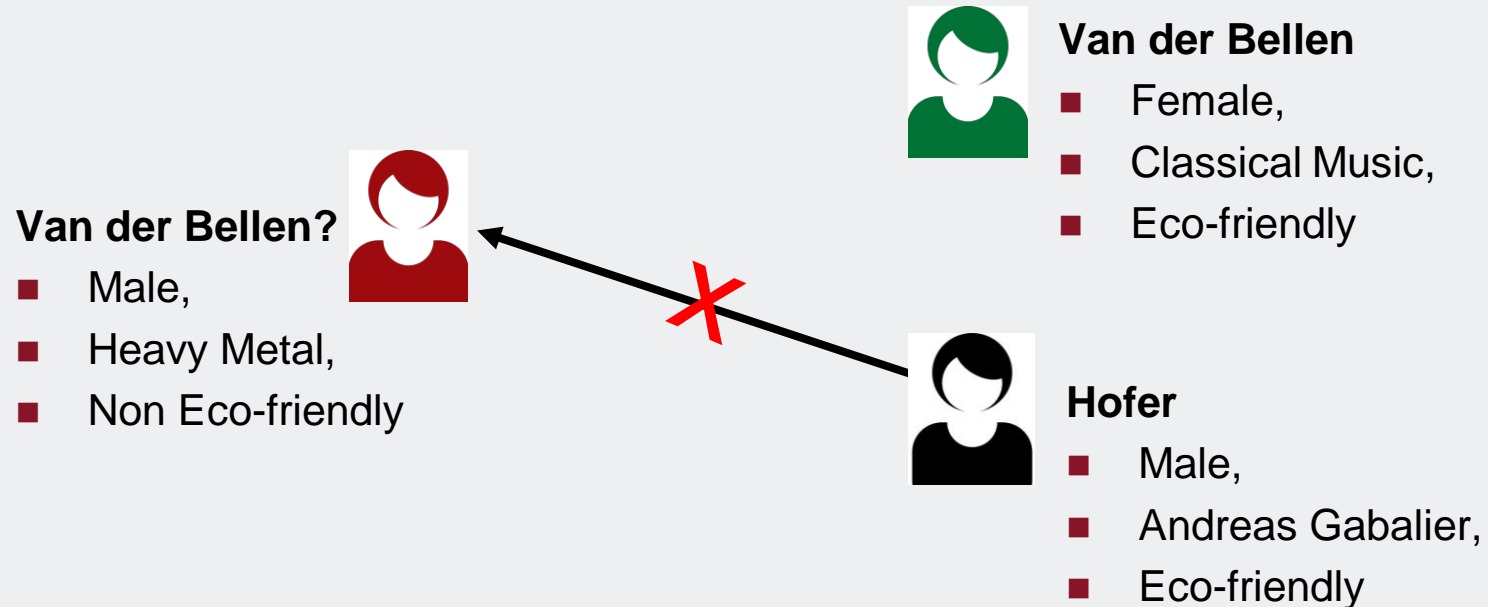
Individual Level

- Independent individuals
- Described by **attributes** (e.g., gender, age, personality)
 - *Which attributes predict political view/voting behavior?*
- Distribution of attributes, correlation between attributes, statistical inference
- Methods: descriptive statistics, t-tests, regression models, etc.

Individual Level (2)

Example: Austrian Presidential Elections 2016

Voting for Van der Bellen?



Which attributes predict voting behavior/political view?

Group Level

- Sets of individuals
- Described by **associations** among categories (e.g., affiliation, music taste)
 - *Which individuals are in which groups? How does it impact their behaviour?*
- Locations of individuals, similarities between individuals or groups, dimensions that structure the space
- **Method:** Geometric Data Analysis (GDA)
 - (Multiple) Correspondence analysis
 - Principal Component Analysis (PCA)

Social Space

Pierre Bourdieu: Social Space and Habitus

- Social life conceptualized within a social space, where everyone is positioned in relation to one another (“objective relations”)

Habitus:

- Position in the social space effects the way a person acts, thinks, dresses, speaks and gesticulates and also comprises a person’s taste
- Can be seen as a physical manifestation of the social world

Geometric Data Analysis (GDA)



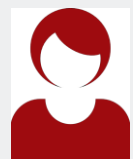
Soccer,
Andreas Gabalier, ...



Golf,
Real Estate, ...



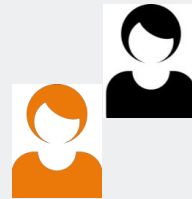
Jazz,
Clubs,
Green Economy, ...



Stock Prizes,
Fancy Food,
Classical Music, ...

Dimensionality reduction:
Principal Component Analysis
(Multiple) Correspondence Analysis

Sports

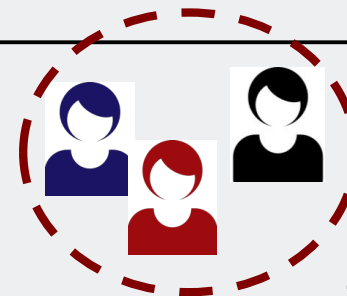


Metric Space



Emerging opinions?

Economy



More likely
to vote for
Van der Bellen?

Music

No interactions!

Network Level

- Dyadic relations between individuals
- Described by **social relations** between individuals (e.g., friendship)
 - *How are individuals connected?*
- Relational patterns, interdependency
- **Methods**: Social network analysis, statistical network models (ERGMs,...)

The Social Networks Perspective

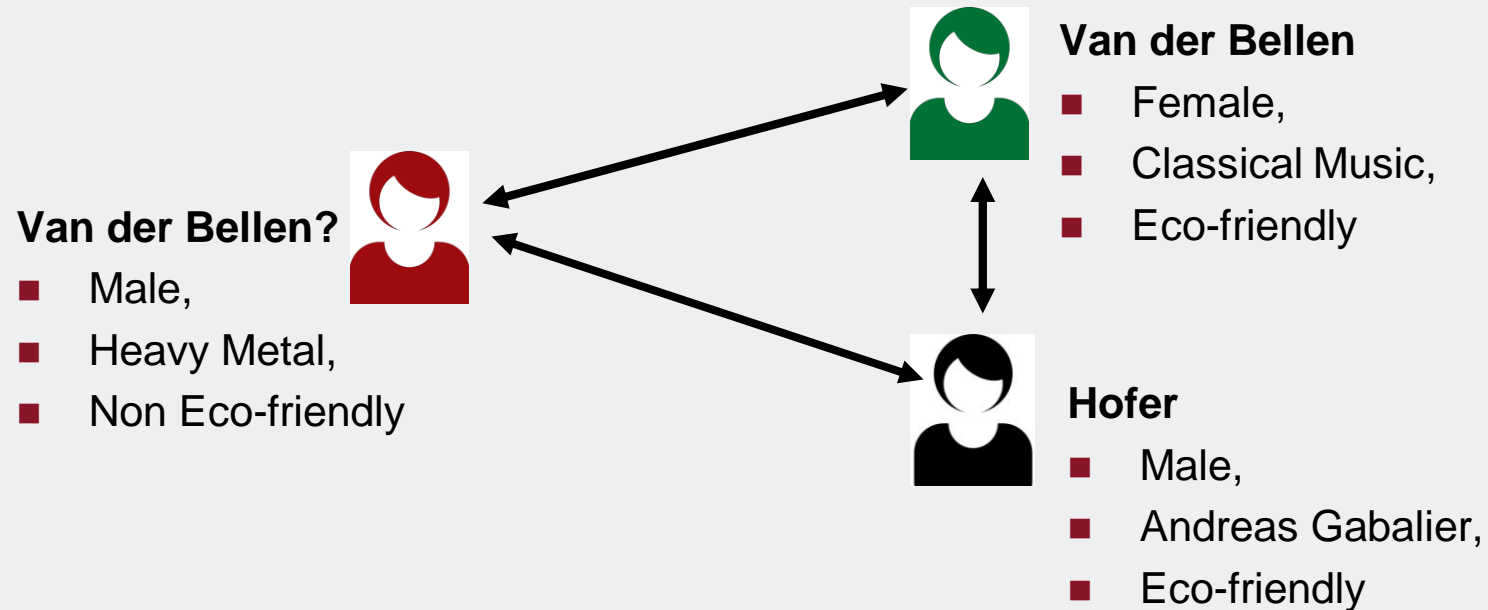
Principles of Social Network Analysis

- Individuals and their actions are considered as interdependent rather than independent;
- Relations between the individuals facilitate the transfer of ideas or behaviors;
- Individual actions are influenced by the position of the individual within the network structure (structure provides opportunities but also constraints them).

Network Level Analysis

Example: Austrian Presidential Elections 2016

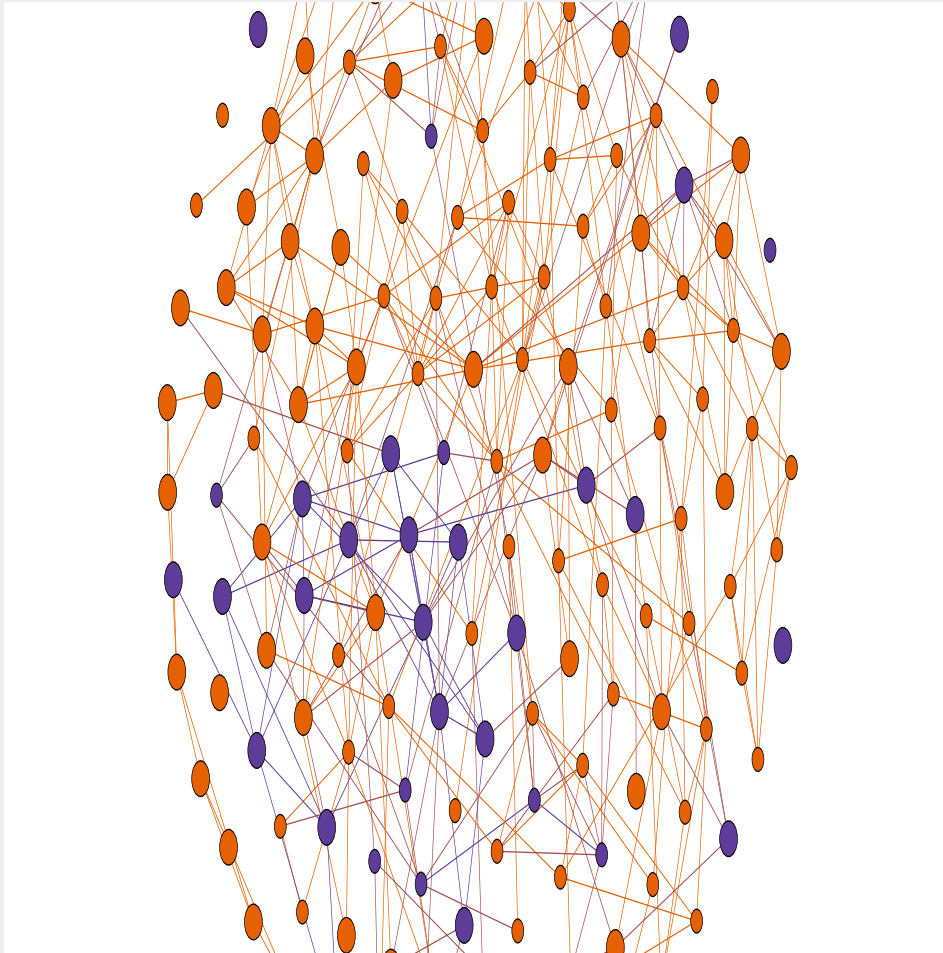
Voting for Van der Bellen?



Our Work

- **Overall goal:** developing new approaches that combine different levels in order to overcome limitations
 - Statistical network models
 - Graph embeddings
- **Online news forum:**
 - Rich and comprehensive dataset (approx. 0.7 TB) comprising
 - User postings plus metadata (since 2000)
 - Postings, user data and votes (millions of anonymous users, hundreds of thousands user accounts, millions of postings and votes)
 - Articles plus metadata (since 2000)
 - Including keywords (millions of articles)
 - Website usage data (since 2012)
 - Clickstreams (billions of page impressions)

Statistical Network Models



Statistical Network Models, e.g., Conditional Random Field Model

- **Outcome variable**: voting for Van der Bellen
- **Predictor variables**: gender, age, music taste, eco-friendly, etc.
- **Social relations**: pairwise communication

Network size: 160

37 **Hofer**

123 **Van der Bellen**

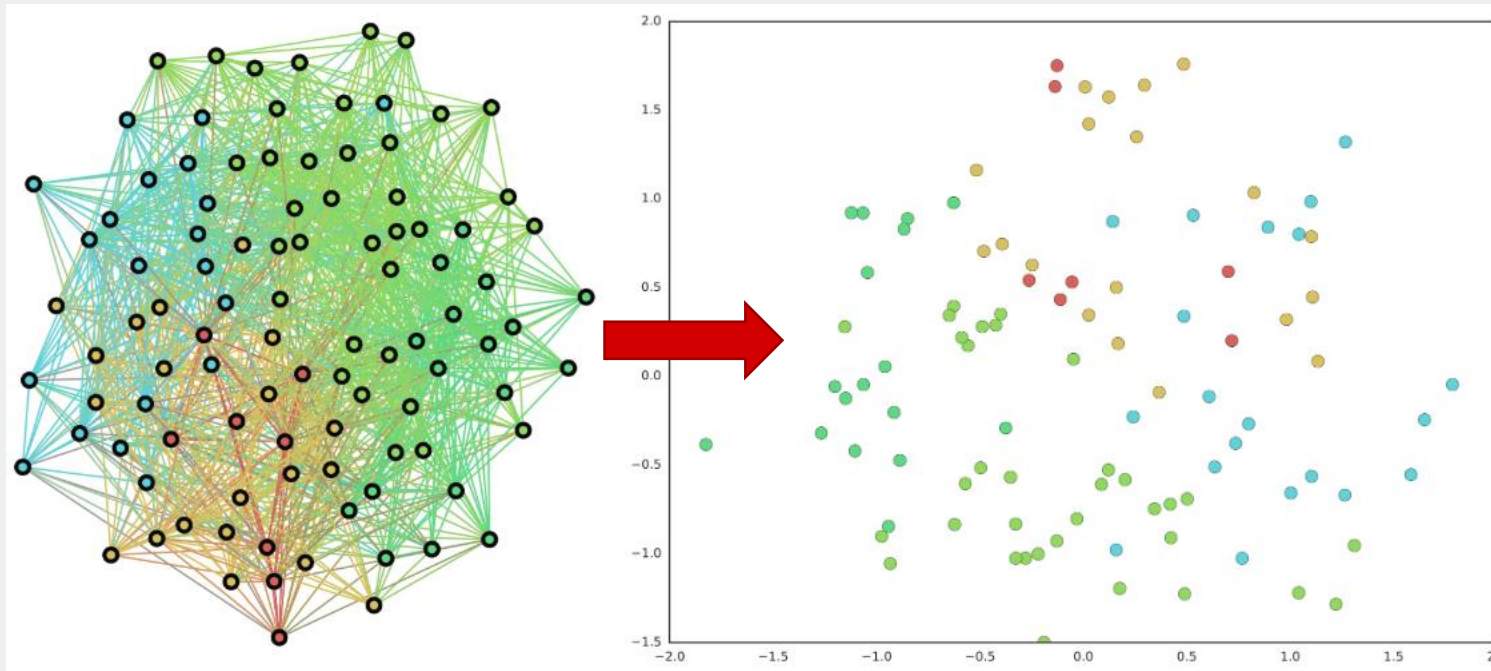
76 female

(bigger nodes)

Graph Embeddings

- **Mapping graphs onto metric spaces:**

- Graph embedding methods (DeepWalk, node2vec,...) learn a continuous vector space for the graph, assigning each node (and/or edge) in the graph to a specific position in a vector space



Studying filter bubbles / polarization

Conclusion

- Overall goal is to model and to understand, how opinions emerge, change and get exchanged
- Traditional theories focus on specific aspects related to different levels
- Today social online communities are generating enormous amounts of data capturing detailed user behavior and their interactions
- In parallel, novel approaches emerge that allow to handle all the data and (maybe) introduce new perspectives
- Provides unprecedented opportunities to study user dynamics and opinion forming and other phenomena over time based on a more comprehensive, multi-level user model
- However, interdisciplinary approaches are required

→ Work published in ***Nature Human Behaviour***

Thank you!