Social Network Analysis in Practice

Projektseminar WiSe 2018/19 Tim Repke, Dr. Ralf Krestel



Social Network Analysis in Practice

Universitär

HPI

Tim Repke 17.10.2018

Un^{iversit}äx HPI

About Me





What You Learn in this Seminar

- Overview of topics in Social Network Analysis
- Deep understanding of one aspect
- Explore heterogeneous data
- How to find, implement, and reproduce research papers
- How to write a research abstract

Social Network Analysis in Practice

Schedule





Social Network Analysis in Practice

Tim Repke 17.10.2018

Schedule





Social Network Analysis in Practice

Schedule (pt. 1) Overview Talks



Date	Topics	
17.10.2018	Introduction	
24.10.2018		
31.10.2018		
07.11.2018	 Introduction to Social Network Data Data acquisition and preparation Data interpretation 	Reformationstag
14.11.2018	Introduction to Measuring Graphs Node/dyad-level metrics Dyad/group-level metrics Interpreting metrics 	
21.11.2018	Introduction to Advanced Analysis Community detection and clustering Mixed membership models Information diffusion Graph/pode embeddings 	Social Network Analysis in Practice
28.11.2018	Introduction to Network Visualisation Matrix-based visualisation Traditional visualisation/graph layout 	тт керке 17.10.2018 7

J

Schedule (pt. 1) Overview Talks – Topic Assignment



- Topics will be assigned ASAP!
- Please send email to <u>tim.repke@hpi.de</u>
 - Subject: [SNAiP] Topic preferences
 - Data Engineering students preferred
 - First-come-first-serve, but please cancel if you change your mind
 - Please send top 3 topics ordered by preference
 - Registration deadline: Friday, 19.10., 18:00

Social Network Analysis in Practice

Schedule (pt. 1) Overview Talks - Presentation Specs

- Length: 25 mins interactive lecture¹
- Presentation should cover all aspects of the topic
- Highlight pros/cons/applications of presented methods
- For general (network) algorithms: make it SNA specific
- Students should get overview of what can be done, not how one specific thing works
- Please coordinate, so that talks in each lecture fit together
- We'll meet to discuss slides the week before your presentation, please send finished slides to <u>tim.repke@hpi.de</u> the day before
 - Feel free to send me a selection of topics/methods you want to cover earlier on!



University.



Schedule





Social Network Analysis in Practice

Schedule (pt. 2) Exploration



Date	Topics
05.12.2018	Lecture Enron Dataset, practical tips working with network data
12.12.2018	
19.12.2018	Presentations Findings of Data Exploration

- Datasets, project topics, tools/libraries will be presented on 05.12.
- Get aquainted with data and tools and prepare small, informal presentation with #datascience stuff
- More on that then...

Note: I will choose and prepare project topics based on **your intestests**! Your requests are very welcome! Social Network Analysis in Practice

Schedule





Schedule (pt. 3) Project



Date	Topics
09.01.2019	Groups work on projects weekly meetings with each group
16.01.2019	
23.01.2019	
30.01.2019	
06.02.2019	Presentations Final Project Result

- Groups of 2-3 students
- Implement a paper, method, idea you like with real-world data
- I'll meet with each group once a week for feedback
- More on that in December!

Social Network Analysis in Practice

Schedule





Social Network Analysis in Practice

Schedule (pt. 4) Paper



Date	Topics
13.02.2019	
20.02.2019	
27.02.2019	Paper Submission Deadline
06.03.2019	Peer Review
13.03.2019	
20.03.2019	"Camera Ready Version" of Paper

- Single-blind review
- 2 pages (+ references)
 - ACM proceedings template¹
 - Each student writes about one aspect of the group project
 - Either individual papers or one longer paper

Social Network Analysis in Practice

Seminar Philosophy





Grading





Social Network Analysis in Practice

Questions so far?





Social Network Analysis in Practice



Introduction to Social Network Analysis

What is a Social Network?





- Set of actors
 - E.g. people
- Set of ties
 - Friendships
 - Follows, likes,...
- Research in sociology, economics, maths, computer science
 - Detect salient structures
 - Extract knowledge, reduce data
 - Analyse social behaviour



What is a Social Network? (cont.)

- Nodes can be individuals or collectives (or both)
- Relations can be
 - Similarities
 - Same location, same time
 - Same attribute (e.g. gender or association)
 - Roles (mother of, friend of, ...)
 - Cognition (knows, likes, ...)
 - Interaction/events (emails, posts, retweets,...)
- Heterogeneous data
 - Attributes at nodes or edges (text, time, location,...)

Social Network Analysis in Practice

Why Social Network Analysis?

- Actors affect each other
- Analyse diffusion
 - Epidemiology (patterns of disease transmission)
 - Information (social influencer, ...)
- Business
 - Marketing, Opportunities, Market segmentation
 - Optimise internal operations
- Investigation
 - Uncover crime and corruption
 - Data-journalism, audit, ...

Social Network Analysis in Practice

Universitäx

örsdam



General - Data acquisition and preparation



- Where does data come from?
- What data to collect?

000

- How to form hypotheses?
- How to limit the collected data?



Social Network Analysis in Practice

Tim Repke 17.10.2018

24

- How to deal with boundaries of collected data?
- How to interpret non-existing links?
- Where do errors come from?

- Data interpreation

General

- How to filter data for analysis?
- How to measure the error of filtering and selecting?



Social Network Analysis in Practice

Universitäx

. sdam



Measuring Graphs - Node/dyad-level metrics

- How can **nodes** be described?
 - Distance to one another
 - "Importance" of a node
 - Connectivity
 - Centrality
- Rank, filter, compare nodes













Measuring Graphs - Dyad/group-level metrics

- How can **relations** or the **entire graph** be described?
 - Within-group connectivity
 - Clustering coefficient
 - Effective aggregation of node measures
- Higher level of analysis
- General insights



Social Network Analysis in Practice

Universitäx

örsdam

Tim Repke 17.10.2018



¹<u>http://konect.cc/networks/enron-rm/</u>

n = 87,273 m = 1,147,126 m = 321,288 I = 13,080 s = 49,130,407 z = 13,799,850,890 x = 3,784,732,758,298 t = 1.178.289 q = 92,655,137 $T_4 = 938,356,410$ d_{max} = 38,778 $d_{max}^{+} = 32,613$ $d_{max}^{-} = 6,165$ d = 26.288 2 $p = 4.234 95 \times 10^{-5}$ m = 3.570 40 N = 84,220 $N_{s} = 9,160$ N^r_s = 0.104 958 δ = 14 $\delta_{0.5} = 4.38396$ $\delta_{0.9} = 5.79471$ δ_M = 5 δ_m = 4.890 80 G = 0.907 479

Measuring Graphs - Interpreting metrics

Statistics

Unique edge count

Size

Volume

Loop count

Claw count

Cross count

Triangle count

Square count

4-Tour count

Maximum degree

Maximum outdegree

Maximum indegree

Average edge multiplicity

Relative size of LSCC

50-Percentile effective diameter

90-Percentile effective diameter

Average degree

Size of LCC

Diameter

Size of LSCC

Median distance

Mean distance

Fill

Wedge count



Social Network Analysis in Practice

Tim Repke 17.10.2018



Advanced Analysis - Community detection & clustering

Keywords:

- Community Detection
- Node Clustering
- Graph Segmentation







Social Network Analysis in Practice

Tim Repke 17.10.2018

• Grivan and Newman. *Community structure in social and biological networks.* PNAS 2002

HPI

Advanced Analysis - Mixed membership models



- Stochastic Block Models
- Mixed Membership Models



Non-overlapping



Overlapping





Social Network Analysis in Practice

- Blei et al. *Mixed Membership Stochastic Blockmodels.* JMLR 2008
- Williams et al. Mixed Method Social Network Analysis. ORM 2015

Gegendarstellung

Fake News

Advanced Analysis - Information diffusion

- How information travels through network
- Link prediction

Social Network Analysis in Practice

Tim Repke 17.10.2018



31

in Practice

Tim Repke

17.10.2018

Advanced Analysis - Graph/node embeddings

- Node similarity
- Node classification
- Community embedding
- Graph embedding

- Tsitsulin et al. VERSE: Versatile Graph Embeddings from Similarity Measures. WWW 2018
- Anton Tsitsulin et al. *NetLSD: Hearing the Shape of a Graph.* KDD 2018
- Grover et al. node2vec: Scalable Feature Learning for Networks. A., J. Leskovec. KDD 2016





Network visualisation - Matrix-based visualisation

- Visualise network adjacency matrix
- How to utilise block models
- Incorporate time aspects
- Colouring points effectively









Social Network Analysis in Practice

Universita.





Network visualisationTraditional visualisation/graph layout

- Fruchterman and Reingold. *Graph drawing by force-directed placement*. Software: Practice and experience 1991
- Jacomy et al. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one 2014



Social Network Analysis in Practice

Universits.

'dam



Network visualisation - Dynamic and advanced visualisation

- Sallaberry et al. Contact trees: Network visualization beyond nodes and edges. PLOS ONE 2016
- Dang et al. CactusTree: A tree drawing approach for hierarchical edge bundling. PACVIS 2017
- Tao et al. HoNVis: Visualizing and Exploring Higher-Order Networks. PACVIS 2017
- Yang et al. Overlapping communities explain coreperiphery publisher of networks. Proc. IEEE 2014
- Hildenbrand et al. Flexible Level-of-Detail Rendering for Large Graphs. GD 2017
- Efrat et al. *MapSets: Visualizing Embedded and Clustered Graphs*. JGAA 2014
- Bach et al. Towards Unambiguous Edge Bundling: Investigating Confluent Drawings for Network Visualization. TVCG 2016
- Cheong-Iao Pang et al. *Creating realistic map-like visualisations: Results from user studies.* JVLC 2017
- Gibson et al. A survey of two-dimensional graph layout techniques for information visualisation. InfoVis 2012



(a) Skupin, 2004b







Tim Repke 17.10.2018

34





Universits.

'dam

I'm here to help!

- Don't hesitate to email or visit me if you need help!
- If something bothers you take action!
 - Constructive criticism warmly welcome
 - Tell me (*explicitly*! I don't understand indirect feedback...)
 - If you don't want to talk to me personally, please communicate anonymously through the FSR

Social Network Analysis in Practice





Fin

Slides and Material online¹



Social Network Analysis in Practice

Tim Repke 17.10.2018

¹<u>https://hpi.de/naumann/teaching/teaching/ws-1819/social-network-analysis-in-practice-ps-master.html</u>

Fin - Remember to send preferences



Date	Topics
17.10.2018	Introduction
24.10.2018	
31.10.2018	
07.11.2018	 Introduction to Social Network Data Data acquisition and preparation Data interpretation
14.11.2018	Introduction to Measuring Graphs Node/dyad-level metrics Dyad/group-level metrics Interpreting metrics
21.11.2018	Introduction to Advanced Analysis Community detection and clustering Mixed membership models Information diffusion Graph/node embeddings
28.11.2018	 Introduction to Network Visualisation Matrix-based visualisation Traditional visualisation/graph layout Dynamic and advanced visualisation



38

References

- https://github.com/MichaelKreil/twitter-analysis/blob/master/docs/slides_social_bots.pdf
- https://www.michael-kreil.de/
- https://onlinelibrary.wiley.com/doi/abs/10.1002/spe.4380211102
- <u>http://faculty.ucr.edu/~hanneman/nettext/</u>
- http://www.kstoolkit.org/social_network_analysis
- http://www.thenetworkthinkers.com/
- <u>http://orgnet.com/stories.html</u>
- http://www.realsystems.eu/?page_id=41
- http://www.alexandramarin.ca/uploads/8/5/7/4/85749/marin_and_wellman 2010 social_network_analysis_an_introduction.pdf
- https://study.sagepub.com/borgatti2e
- http://snap.stanford.edu
- http://igraph.org/python/
- http://networkx.github.io/

Social Network Analysis in Practice



References



- Social Network Analysis with Applications by Ian McCulloh, Helen Armstrong
- Social Media Mining: An Introduction by Reza Zafarani, Mohammad Ali Abbasi
- Computational Network Science: An Algorithmic by Henry Hexmoor
- An Introduction to Exponential Random Graph Modeling by Jenine K. Harris
- Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications by Dean Lusher, Johan Koskinen
- Network Role Mining and Analysis by Derek Doran
- Social Network Analysis: Methods and Applications by Stanley Wasserman, Katherine Faust
- Analyzing Social Networks by Stephen P Borgatti, Martin G Everett
- Grundlagen der Netzwerkforschung: Situation, Mikronetzwerke und Kultur by Christian Stegbauer
- Mining Social Networks and Security Informatics by Tansel Özyer, Zeki Erdem
- Statistical Analysis of Network Data: Methods and Models by Eric D. Kolaczyk
- Networks, Crowds, and Markets: Reasoning about a Highly Connected World by David Easley, Jon Kleinberg
- Mining the Social Web by Matthew A. Russell

Social Network Analysis in Practice