Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Mining Social Graphs

Text, Web and Social Media Analytics Lab

Prof. Dr. Diana Hristova

Exercise 7: Social Media

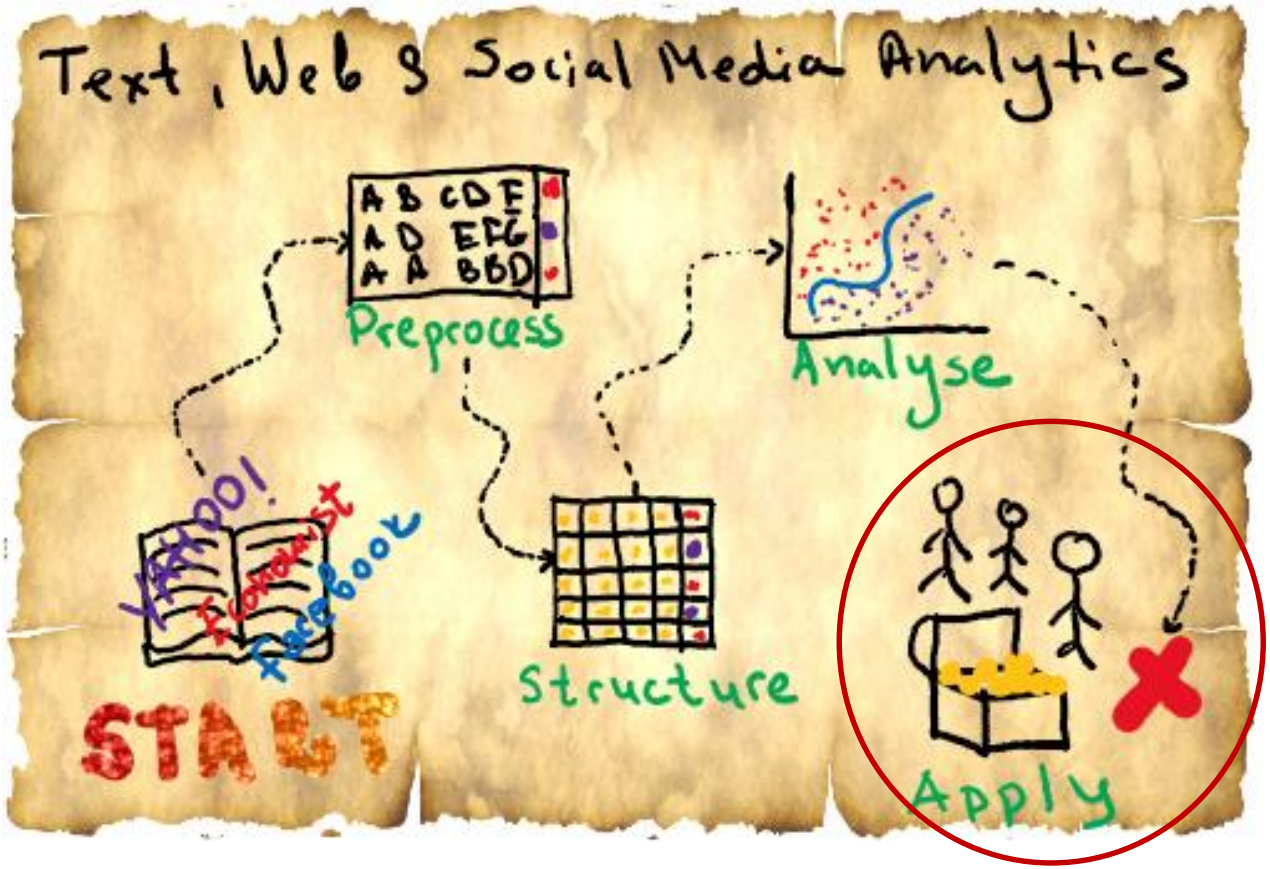
Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Can one group present please?

What did we learn last week?

Social Media Analytics: Treasury map

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law



Course structure

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Date	Lecture	Exercise
12.04.2021	Introduction	Technical Installation
19.04.2021	Text Preprocessing	Projects kick-off
26.04.2021	Text Representation	Preprocessing Newsgroups
03.05.2021	Text Representation (2)	Text Representation Newsgroups
10.05.2021	Text Classification	Text Representation Newsgroups (2)
17.05.2021	Text Clustering/Capgemni	Newsgroups Topic Classification
31.05.2021	Text Mining in Social Media	Newsgroups Topic Clustering
07.06.2021	Mining Social Graphs	Sentiment Analysis and Time Series in Twitter
14.06.2021	Projects Status Update	Projects Status Update
21.06.2021	Web Analytics	Mining Social Graphs in Twitter
28.06.2021	Mock Exam	Web Analytics in E-commerce
05.07.2021	Final Presentation	Final Presentation
19.07.2021	Submit Code & Written report	
t.b.a.	Exam	

Projects Status Update

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- Each group should discuss the following points:
  1. What is the task?
  2. Why does it generate added value?
  3. How does the data look like (high-level)?
  4. What is the planned approach?
  5. First results
  6. Issues
- Teams with projects with companies should send their slides for approval to the corresponding company representative in advance.
- Please plan with 20 minutes for your presentation including Q&A**.**

Projects Status Update: Preliminary Schedule

- Group 1 (Signavio): 08:05-08:25
- Group 2 (Signavio): 08:30-08:50
- Group 3 (Strayz): 08:55-09:15

Break: 10 Minutes

- Group 4 (Aam): 09:25-09:45
- Group 5 (BayLev): 09:50-10:10
- Group 6 (Zalando): 10:15-10:35

What will we learn today?

At the end of this lecture, you will:

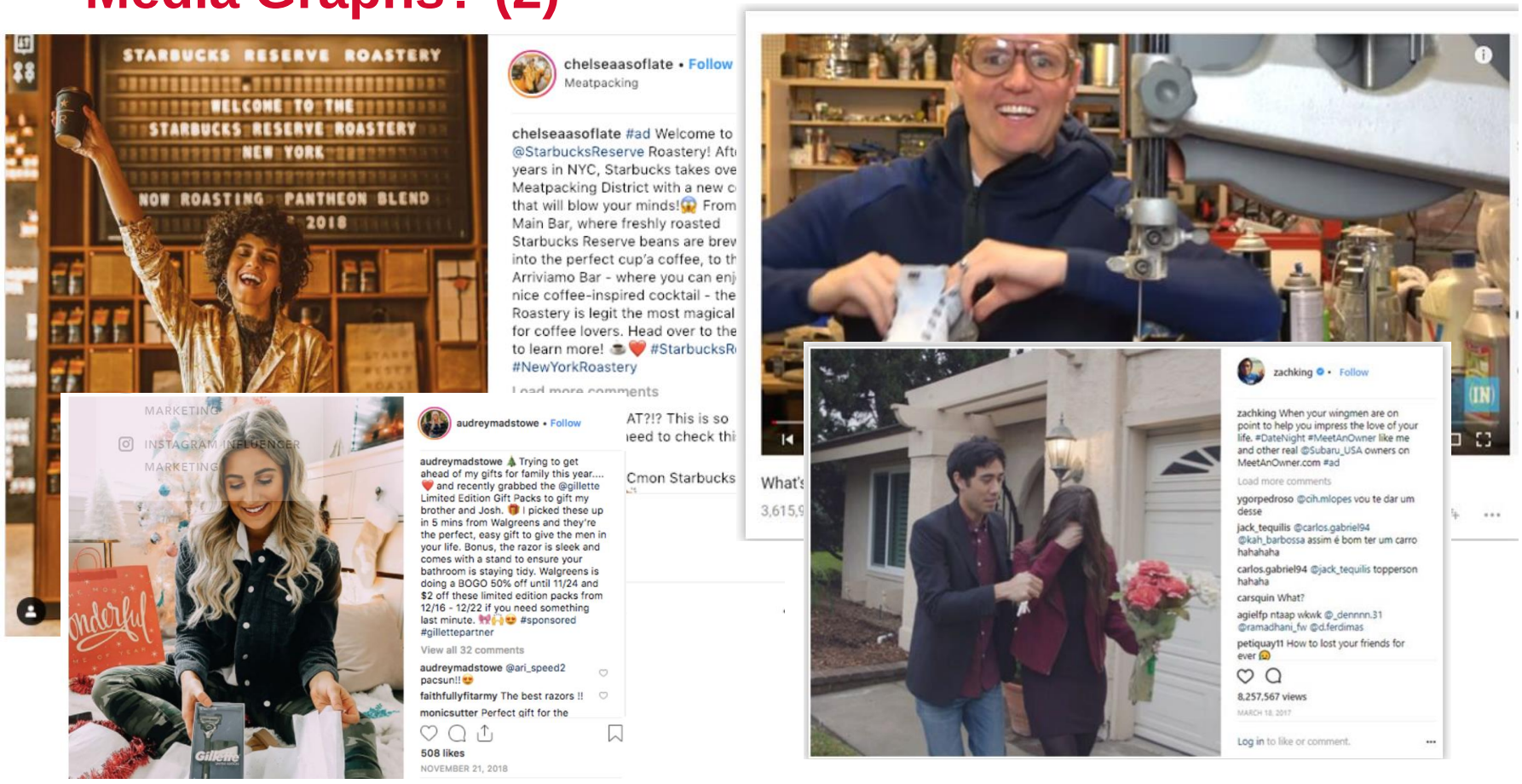
1. Know the motivation behind analysing social media graphs.
2. Be able to represent social media graphs as networks and calculate the following network metrics:
   - Degree centrality
   - Betweenness centrality

Motivation: Why Analyse Social Media Graphs?

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Which politician is likely to influence more people on Twitter?

Motivation: Why Analyse Social Media Graphs? (2)



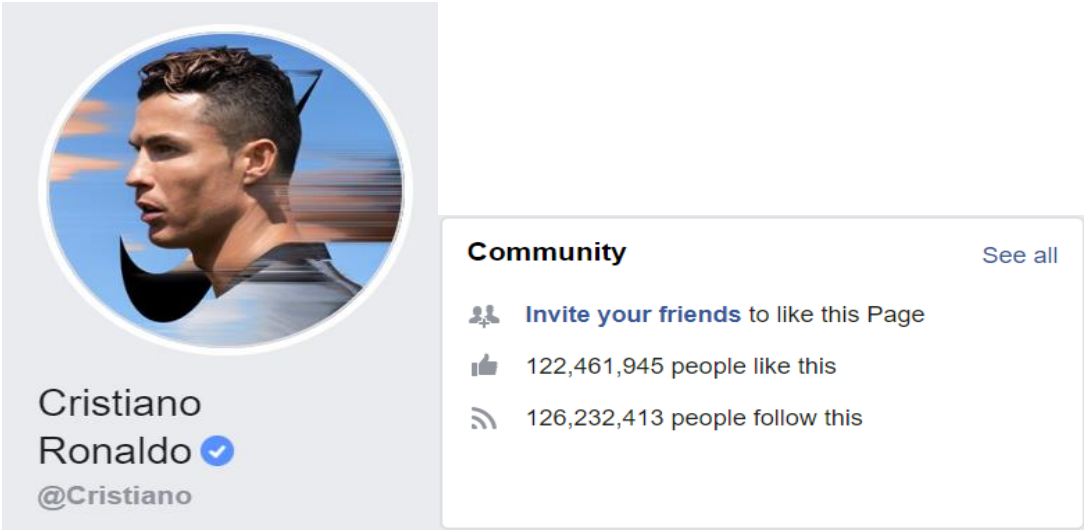
? How to choose your social media partners?

Motivation: Why Analyse Social Media Graphs? (3)

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- Analysing social media graphs can help you identify important users in the network
- Two types of important users in social media graphs we consider here:

1. Users with many direct connections i.e., followers, retweets, likes, comments➔ Use to reach many customers directly e.g. by advertising with them

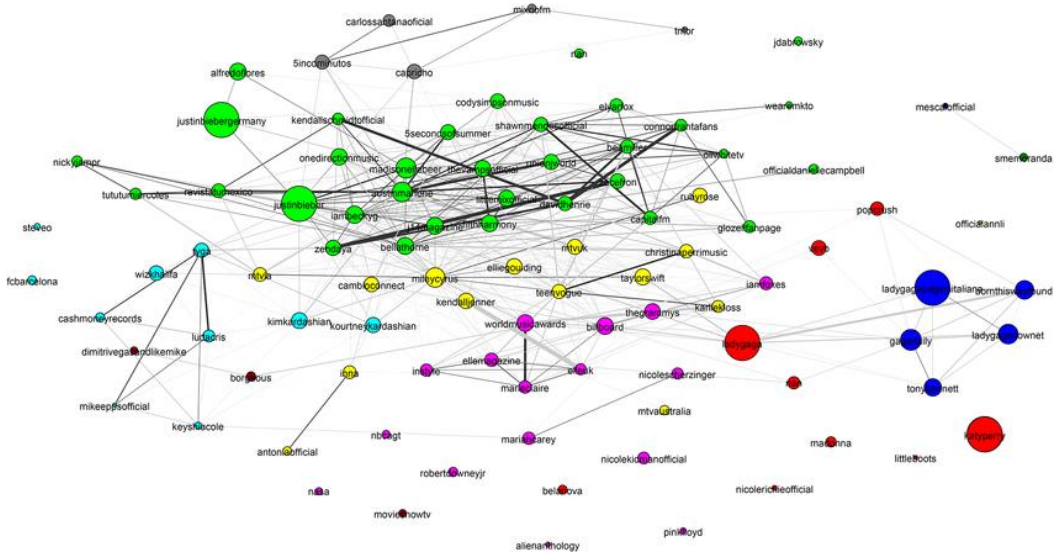


Motivation: Why Analyse Social Media Graphs? (4)

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Two types of important users in social media graphs:

2. Users who are central nodes for information spreading i.e. they transfer ideas between multiple groups (brokers) ➔ use them to reach new customer groups with different interests (e.g. not football fans)

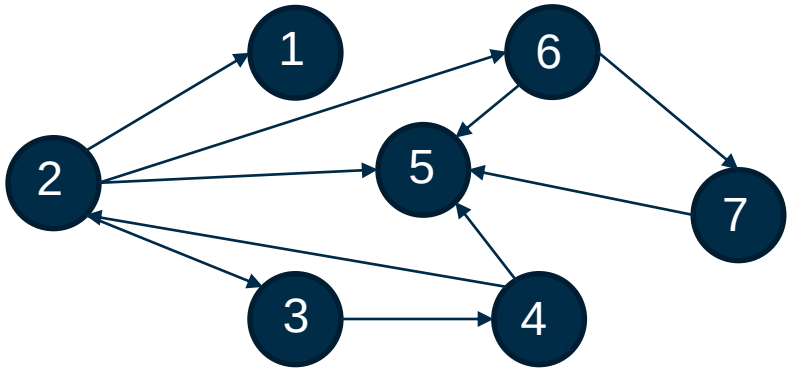


Chamberlain BP, Levy-Kramer J, Humby C, Deisenroth MP (2018) Real-time community detection in full social networks on a laptop. PLoS ONE 13(1): e0188702. https://doi.org/10.1371/journal.pone.0188702

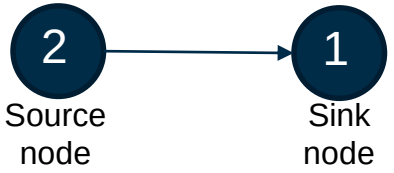
Analysing Social Media Graphs: Network Analysis

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- In order to determine the most important users, **network analysis** is applied.
- A network consists of a set of **nodes** (e.g. users, tweets) and **edges** that connect them.
- An edge shows a relationship between two nodes (e.g. comment, retweet, following, friends).



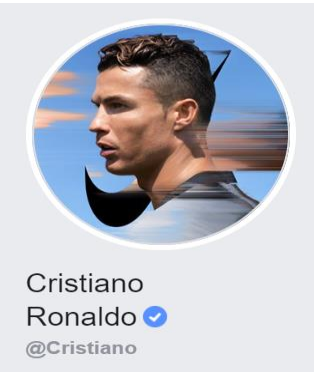
- An edge can be directed (i.e. relationship **from** a source node **to** a sink node) or undirected (i.e. mutual relationship **between** nodes).
- **Focus here:** directed edges and therefore directed networks



Analysing Social Media Graphs: Network Metrics

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

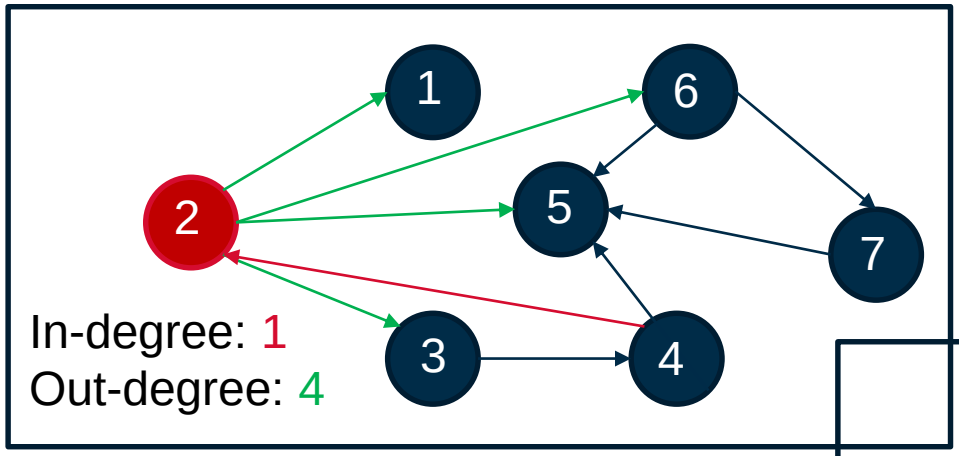
- There a different metrics to measure the role of a node (user/tweet) in a network.
- **Centrality:** measures a node‘s “importance” in the network
- Two types of centrality in this lecture: **degree centrality** and **betweenness centrality**
- **Degree centrality:** number of edges connected to a node i.e. “…the ability of a node to communicate directly with others.” (Estrada Knight 2015, p. 143)
  - In-degree: how many edges go **into** a node (e.g. how often someone retweets)
  - Out-degree: how many edges go **out of** a node (e.g. how often is someone retweeted)

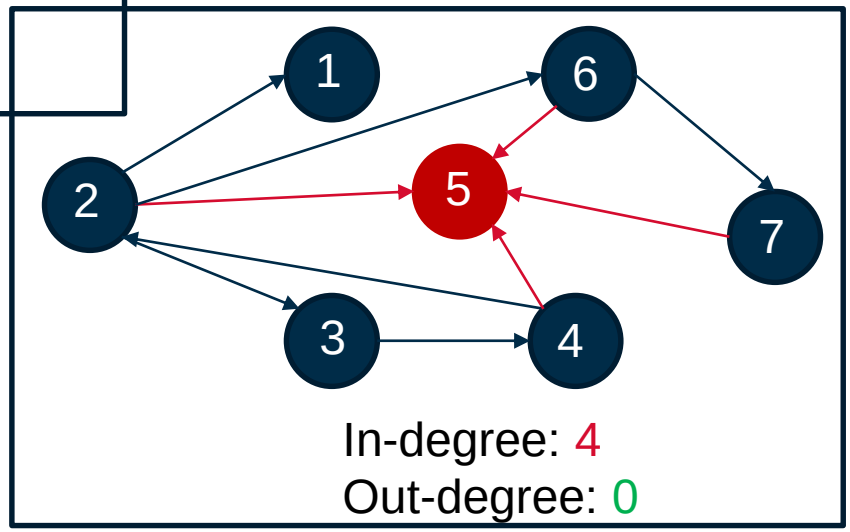


Ernesto Estrada and Philip A. Knight (2015): A First Course in Network Theory, Oxford University Press, 272 pp., £29.99, ISBN 9780198726463. In: *Statistical Papers* 58 (4), S. 1283–1284. DOI: 10.1007/s00362-017-0961-1.

Network Metrics: Degree Centrality

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law



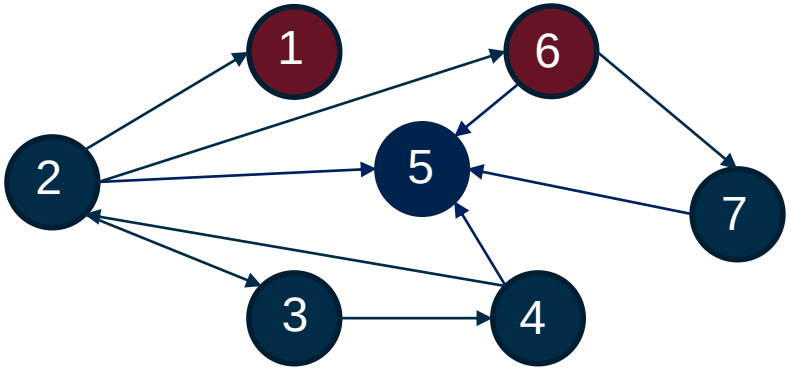


- ➔ User 2 is being often retweeted.
- ➔ User 5 retweets very often.

Network Metrics: Degree centrality Exercise

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

What is the in- and out-degree of the below nodes?

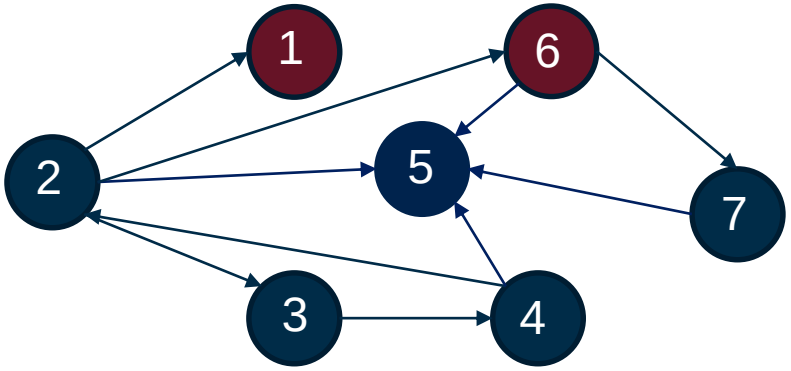


Node	In-degree	Out-degree
1		
6		

Network Metrics: Degree centrality Exercise

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

What is the in- and out-degree of the below nodes?



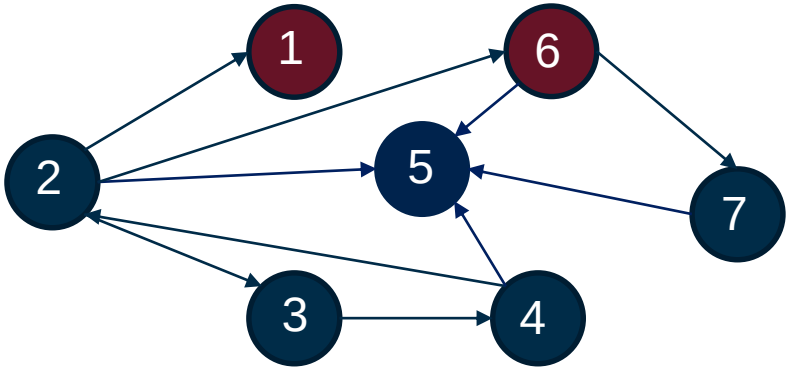
Node	In-degree	Out-degree
1	1	0
6	1	2

! **Note:** degree centrality is often normalised by dividing over (n-1) with n=number of nodes

Network Metrics: Degree centrality Exercise

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

What is the in- and out-degree of the below nodes?



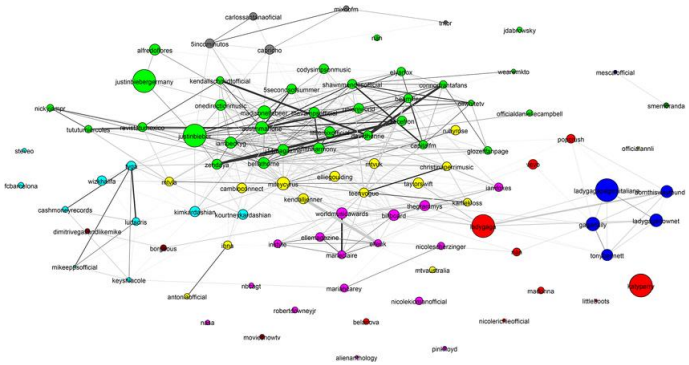
Node	In-degree	Out-degree
1	1	0
6	1	2

? What is a disadvantage of degree centrality?

Network Metrics: Betweenness Centrality

- **Degree centrality** focuses on the user‘s importance, but does not consider the **flow of information** to other users who are not directly related to her. ➔ Betweenness centrality

- **Betweenness centrality:** „… a measure of the influence of a vertex over the **flow of information** between every pair of vertices under the assumption that information primarily flows over the **shortest paths** between them.” (p. 1)[1]

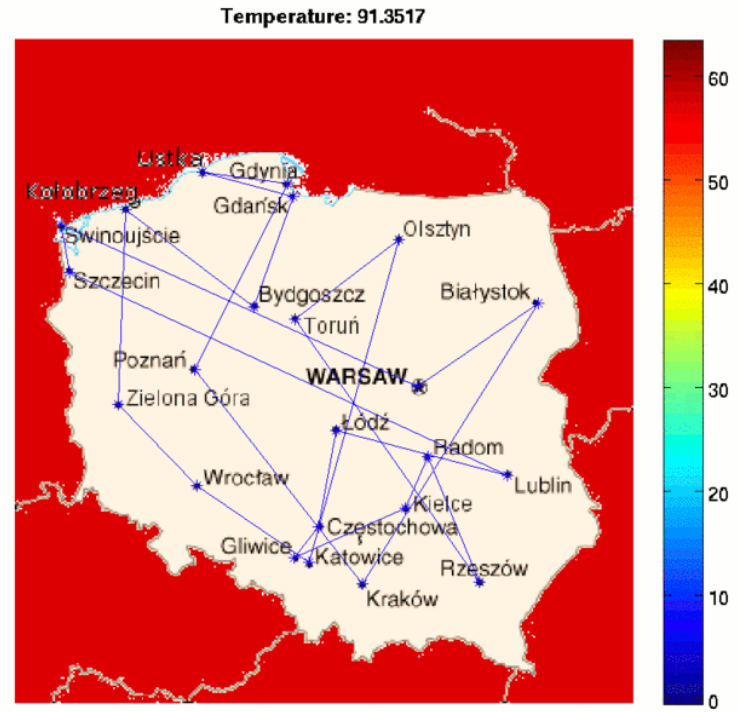


- **Shortest path:** what is the quickest connection (e.g. the least number of users) between two nodes? ➔ we assume information is spread efficiently

[1]Unnithan, Raghavan, et al. "Betweenness centrality in some classes of graphs." *International Journal of Combinatorics* 2014 (2014).

Network Metrics: Betweenness Centrality (2)

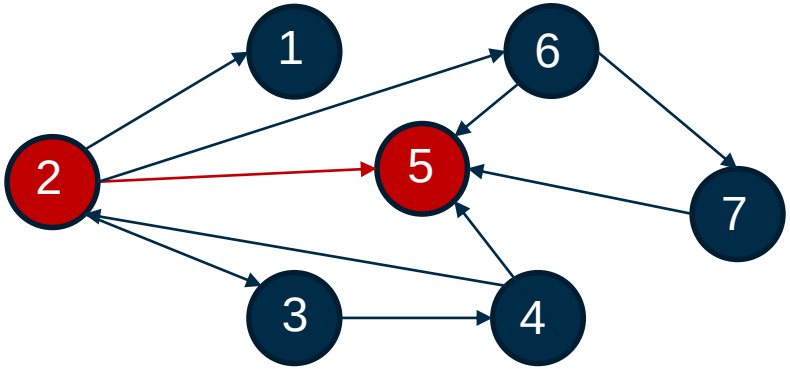
- Betweenness centrality is based on **transportation theory.**



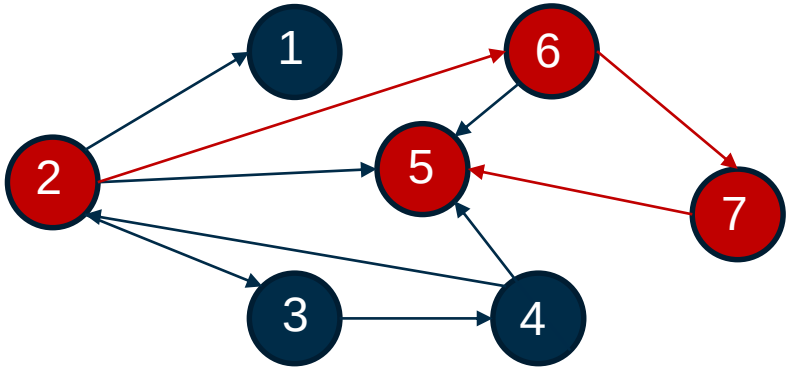
https://commons.wikimedia.org/wiki/File:Sa_poland_tsp.gif

Network Metrics: Betweenness Centrality (3)

Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law



Shortest path from 2 to 5

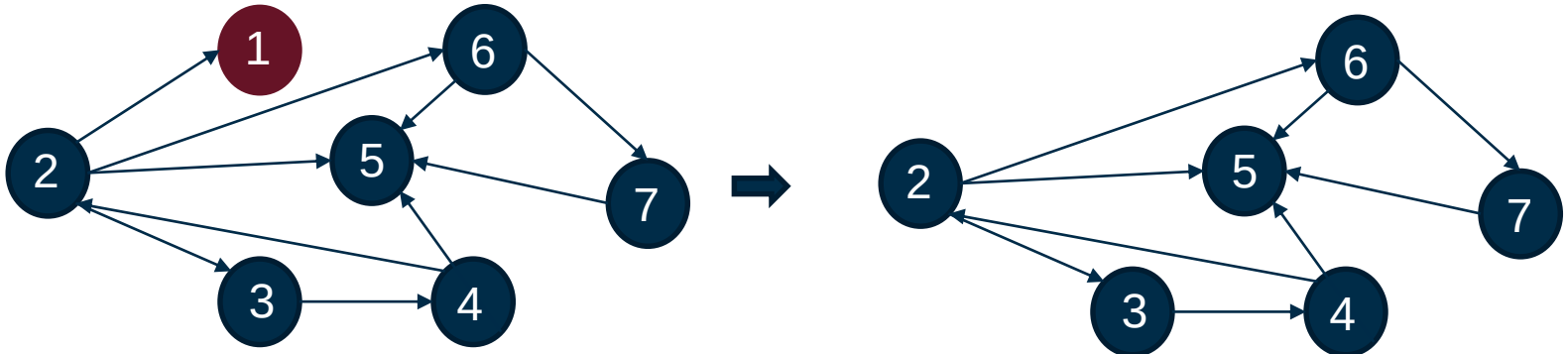


Not the shortest path from 2 to 5

Network Metrics: Betweenness Centrality Calculation

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- **Betweenness centrality:** „… a measure of the influence of a vertex over the **flow of information** between every pair of vertices under the assumption that information primarily flows over the **shortest paths** between them.” (p. 1)[1]
- **Betweenness centrality calculation:** for a given node, determine how many times it will interrupt the shortest-path flow of information in the network, if removed.
- **Example:** remove Node 1

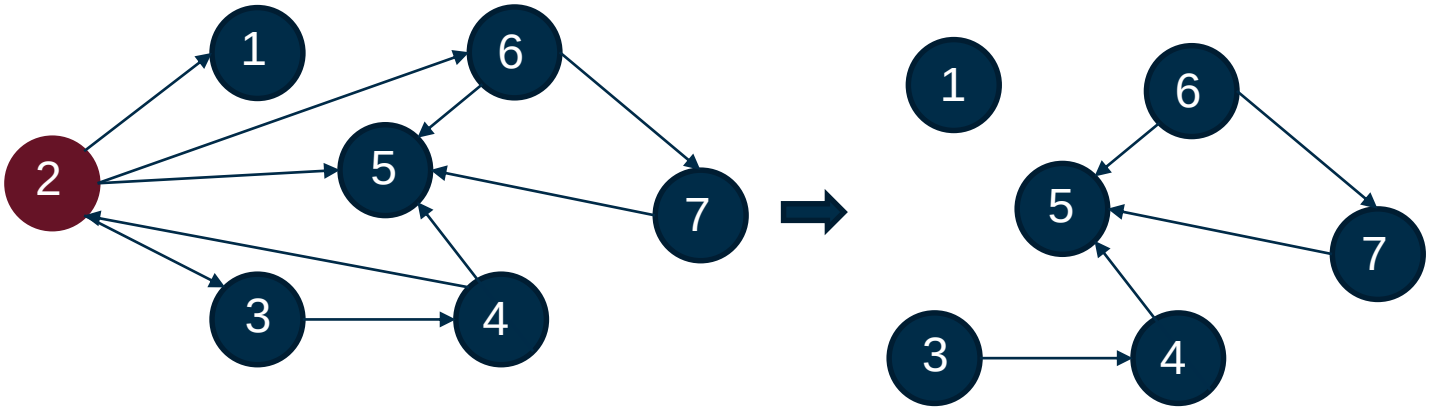


- Node 1 interrupts the shortest path zero times

➔ removing it is not so crucial for the information flow in the network

Network Metrics: Betweenness Centrality Calculation (2)

- **Example:** remove Node 2



- Node 2 interrupts the shortest path from:
  - Node 3 to Node 1, 6 and 7
  - Node 4 to Node 1, 3, 6, and 7

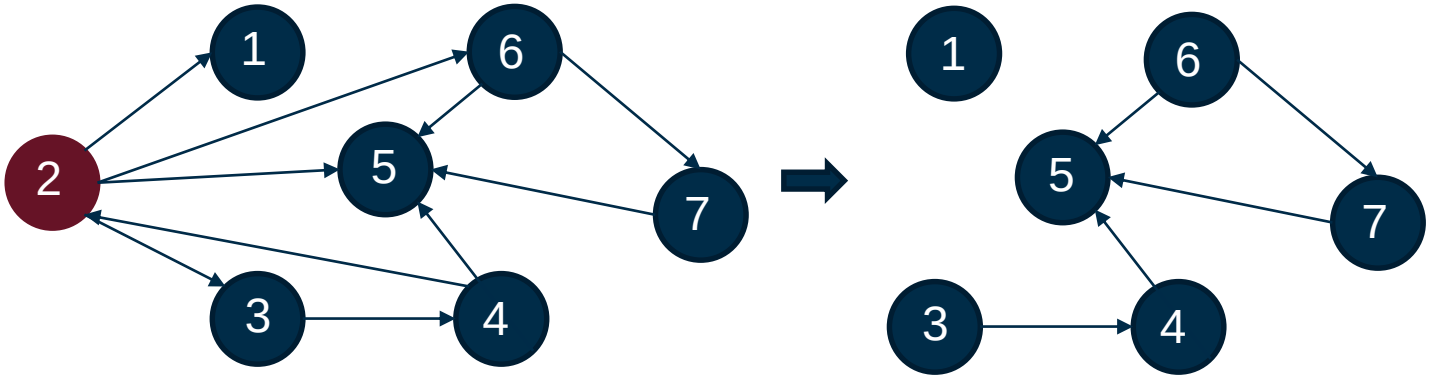
➔ 7 times

➔ Removing it is very important for the flow of information.

Network Metrics: Betweenness Centrality Calculation (3)

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- In order to be able to compare the betweenness centrality of different networks, it is normalised by (n-1)(n-2), where n is the number of nodes in the network (here 7).



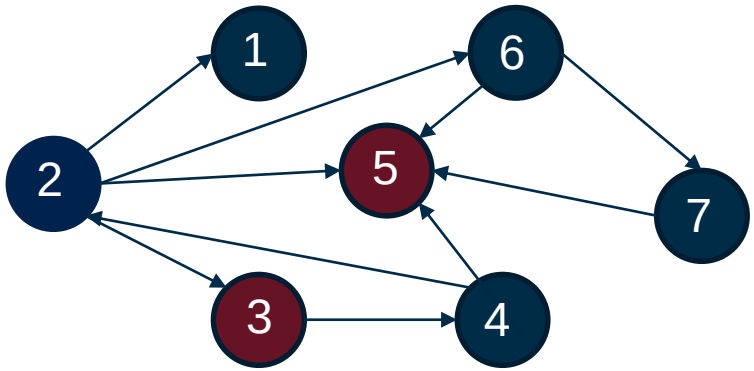
- Node 2 interrupts the shortest path 7 times➔ normalised betweenness centrality =7/30

Network Metrics: Betweenness Centrality Exercise

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- **What is the normalised and the not normalised betweenness of the below nodes?**

Node	Betweenness (not normalised)	Betweenness (normalised)
3		
5		

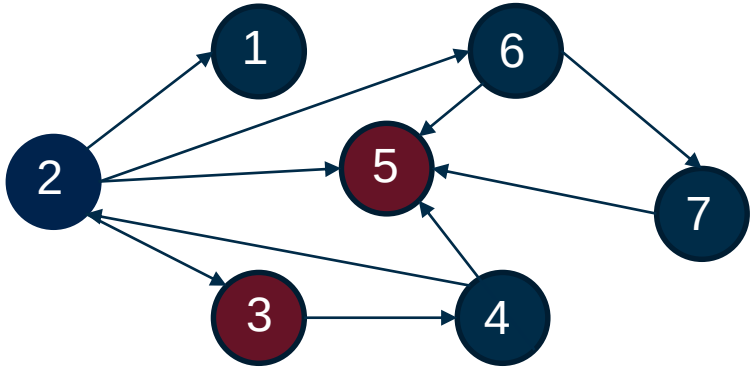


Network Metrics: Betweenness Centrality Exercise

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

- **What is the normalised and the not normalised betweenness of the below nodes?**

Node	Betweenness (not normalised)	Betweenness (normalised)
3	1	0.03
5	0	0



Summary and Outlook

Summary:

- Mining social graphs is important because it helps identify important users and their influence on the flow of information.
- Social graphs can be modelled with network analysis tools and analysed with network metrics.
- Degree centrality is determined the number of direct connections of a user.
- Betweenness centrality determines the importance of the user for the flow of information in the network. It can be normalized for comparison between networks.

Outlook: Social graphs' analytics can be combined with web analytics to determine optimal marketing strategies.

Questions?

Exercise 8

Hochschule für Wirtschaft und Recht Berlin
Berlin School of Economics and Law

In a minute, six break-out rooms will be created. Choose the room that corresponds to your group in Moodle e.g. Room 1= Group 1. In your project group discuss and document the solution for Exercise 8 (in Moodle).