

Wer mit wem?

# Netzwerkanalyse im Datenjournalismus

Prof. Dr. Katharina Zweig

# „Network science“



Ein Netzwerk (in der Analyse komplexer Netzwerke) ist ein **Modell** der **Beziehungsstruktur** von einer **definierten Menge an Personen (Institutionen, etc.)**.

Fast immer nur **eine Beziehung**, vielleicht auch einmal zwei. Damit nur ein Ausschnitt der Wirklichkeit.

Führt **quantitative Analyse** von **indirekten Effekten** durch: wer beeinflusst wen, wer finanziert wen, wer infiziert wen?

Dazu muss es einen **Prozess** geben, der das Netzwerk benutzt, und indirekte Effekte erzeugt.



# Zwei Arten von Stories

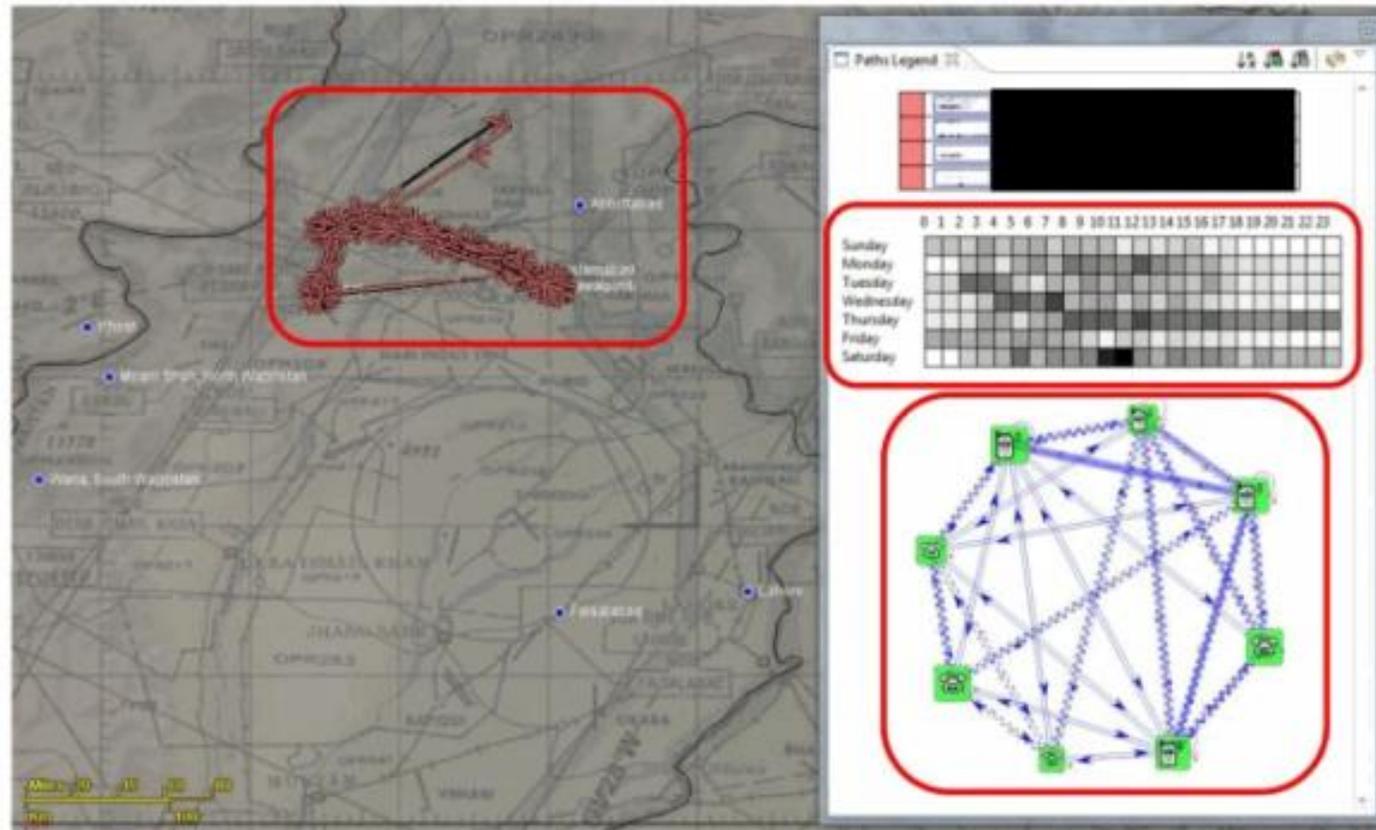
Schlechte Netzwerkanalyse von anderen

# Terroristen fangen mit Netzwerkanalyse



TOP SECRET//COMINT//REL TO USA, FVEY

From GSM metadata, we can measure aspects of each selector's **pattern-of-life**, **social network**, and **travel behavior**



# Terroristenidentifikation SKYNET



TOP SECRET//COMINT//REL TO USA, FVEY

**We've been experimenting with several error metrics on both small and large test sets**

Training Data	Classifier	Features	100k Test Selectors		55M Test Selectors	
			False Alarm Rate at 50% Miss Rate	Mean Reciprocal Rank	Tasked Selectors in Top 500	Tasked Selectors in Top 100
None	Random	None	50%	1/23k (simulated)	0.64 (active/Pak)	0.13 (active/Pak)
Known Couriers	Centroid	All	20%	1/18k		
		Outgoing	43%	1/27k		
+ Anchory Selectors	Random Forest		0.008%	1/14	5	1
				21	6	

Random Forest trained on Known Couriers + Anchory Selectors:

- 0.008% false alarm rate at 50% miss rate
- 46x improvement over random performance when evaluating its tasked precision at 100

Windows  
Wechseln  
aktivieren

TOP SECRET//COMINT//REL TO USA, FVEY

<https://theintercept.com/document/2015/05/08/skynet-courier/>

<https://theintercept.com/2015/05/08/u-s-government-designated-prominent-al-jazeera-journalist-al-qaeda-member-put-watch-list/>

# Top-“Kurier“ der Terroristen laut Algorithmus ist...



TOP SECRET//COMINT//REL TO USA, FVEY

### The highest scoring selector that traveled to Peshawar and Lahore is PROB AHMED Z Aidan

The image displays a map of the region between Peshawar and Lahore, Pakistan, with a purple path connecting the two cities. A profile window for 'PROB AHMED MUWAFAK ZAIDAN' is overlaid on the map. The profile includes a photo and a list of attributes:

- TIDE Person Number: [REDACTED]
- MEMBER OF THE [REDACTED]
- MEMBER OF THE [REDACTED]
- WORKS FOR AL JAZEERA

Windows  
Websel

# Spielkampsche Regel



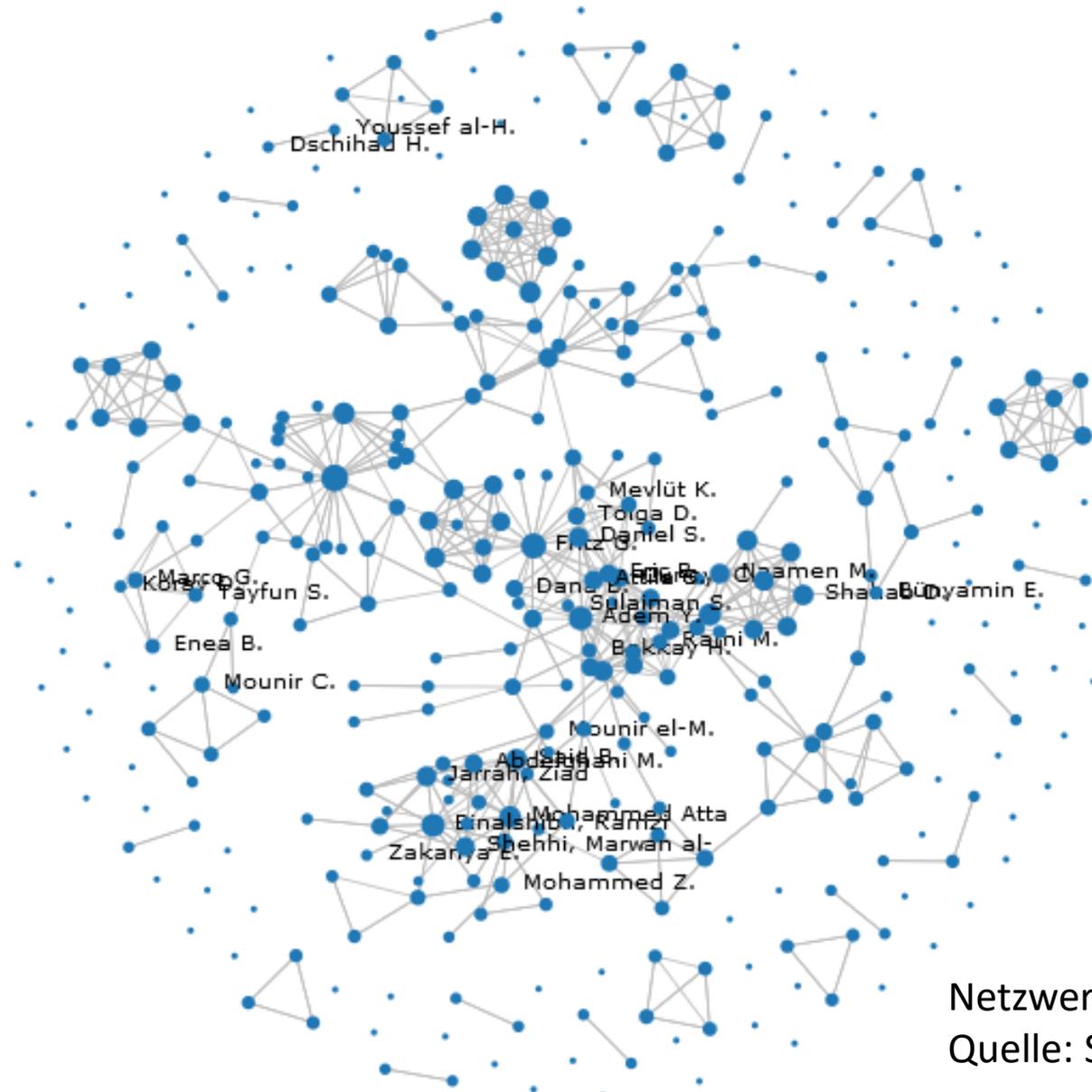
**Alle Algorithmen sind objektiv  
Bis auf die von Menschen gemachten!**



# Zwei Arten von Stories

Selber Netzwerkanalyse betreiben als „Hypothesenfinder“

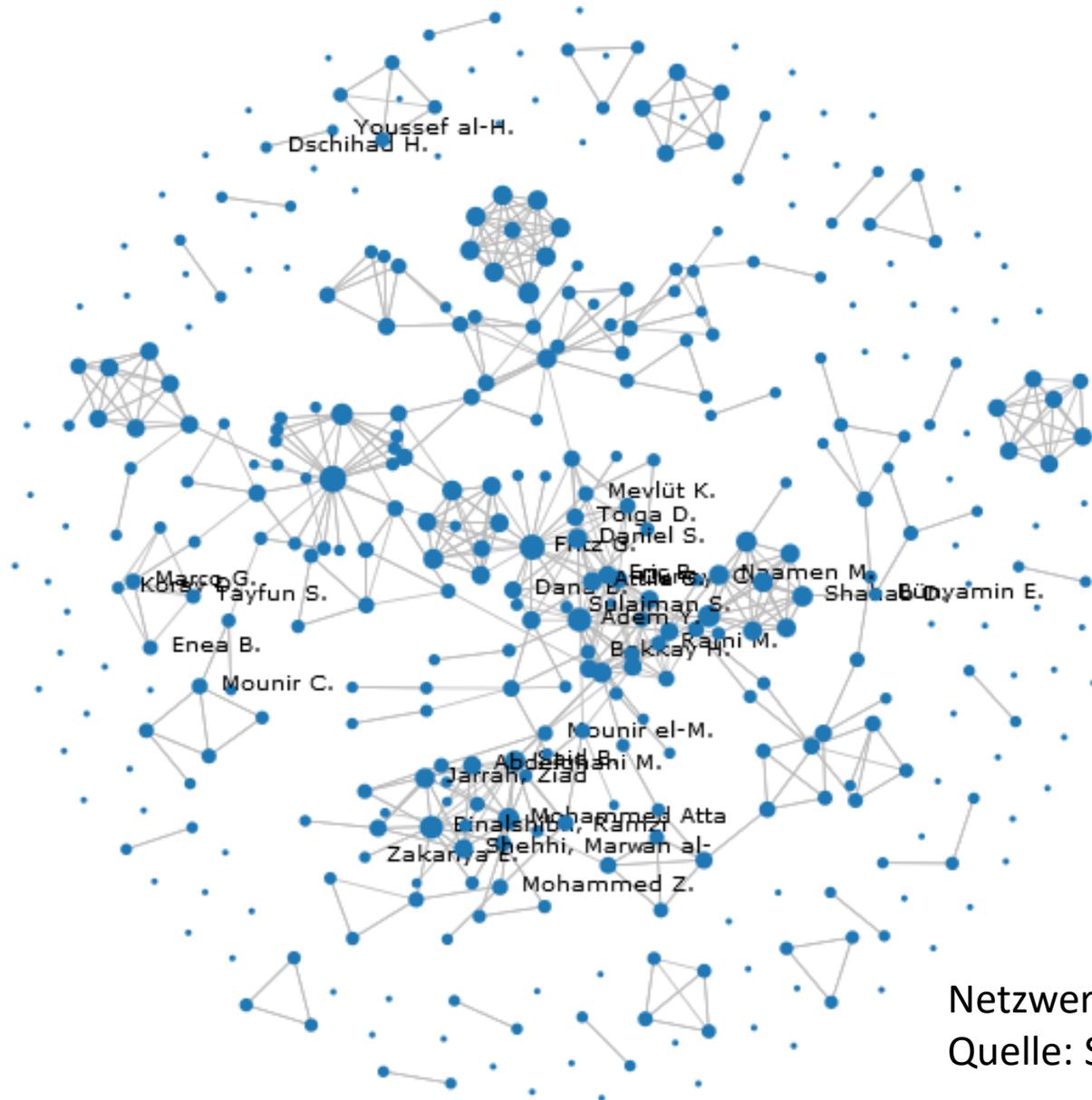
# Können Netzwerke schaden?



Netzwerk mutmaßlicher Terroristen  
Quelle: Spiegel Online

# Können Netzwerke schaden?

- Netzwerke sind visuell nicht „neutral“ (s. Vortrag Lena Groeger)
- Sie verleiten zur Anwendung von Zentralitätsmaßen/Clustering
- Gerade hier schwierig, weil verschiedene Beziehungen gezeigt werden
- Netzwerk nur ein Ausschnitt ist



SPON hat tatsächlich auch keine Netzwerkanalyse betrieben.

Netzwerk mutmaßlicher Terroristen  
Quelle: Spiegel Online

# Fazit



- Netzwerkanalyse modelliert Beziehungen zwischen Dingen, zwischen Personen oder zwischen Dingen und Personen.
- Es geht um die quantitative Analyse indirekter Effekte, verursacht durch Prozesse auf der Netzwerkstruktur, z.B.
  - Informationsweitergabe, Krankheit, biologische Signale, etc.
- Wenn kein Prozess das Netzwerk nutzt, bitte nicht als Netzwerk darstellen.
- Netzwerke sind immer nur Ausschnitte und Modelle der Wirklichkeit mit den üblichen Problemen
  - Fehlerhafte Daten
  - Nicht alle Daten enthalten
  - Aggregiert über zu lange Zeiträume
  - ...
- Relationale Daten sind viel „heißer“ als rein personenbezogene! Nie **UND** nichts mit Klarnamen veröffentlichen.

Richtig angewendet....



.... kann sie weit mehr als nur Trivia hervorlocken....

# Ein bisschen Angeberei



- Identifikation tumorhemmender Biomoleküle
- Analyse von Einkaufskorbdaten
- Bestimmung von hochsignifikant ähnlichen Dokumenten
- Aufreinigung von hochverrauschten (Bio-)Daten
- Analyse sozialer Netzwerke
- Vorhersage von Beziehungen zwischen zwei Nicht(!)-Mitgliedern einer sozialen Netzwerkplattform
- Analyse komplexen Problemlöseverhaltens
- Analyse ökologischer Nischen und Symbiosen
- ...

*(alles eigene Projekte und echt wahr – ansonsten kann Netzwerkanalyse das Weltklima retten, Terroristen identifizieren, die bösen Kapitalisten von den guten trennen, Obamas Wiederwahl garantieren, und und und)*



# Kooperationen erwünscht!

Anfragen an

[zweig@cs.uni-kl.de](mailto:zweig@cs.uni-kl.de)