

ECPR Methods Summer School: Automated Collection of Web and Social Data

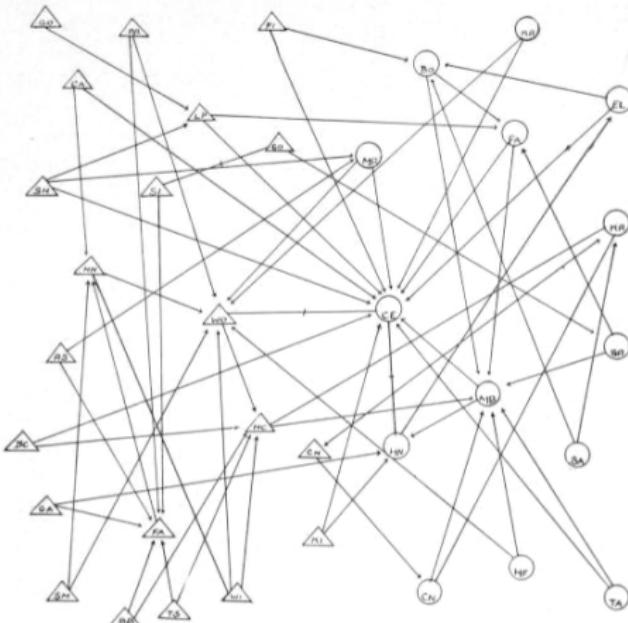
Pablo Barberá

School of International Relations
University of Southern California
pablobarbera.com

Networked Democracy Lab
www.netdem.org

Course website:
github.com/pablobarbera/ECPR-SC103

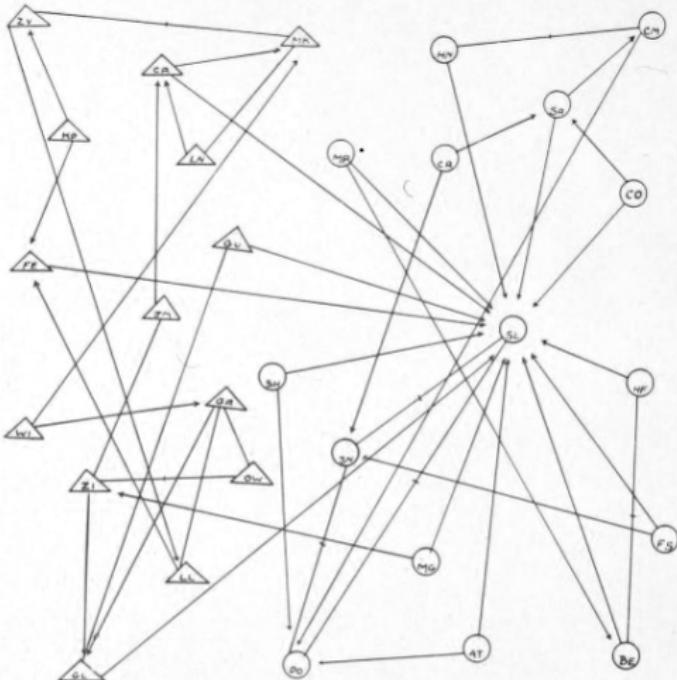
EVOLUTION OF GROUPS



CLASS STRUCTURE, 1ST GRADE

21 boys and 14 girls. *Unchosen*, 18, GO, PR, CA, SH, FI, RS, DC, GA, SM, BB, TS, WI, KI, TA, HF, SA, SR, KR; *Pairs*, 3, EI-GO, WO-CE, CE-HN; *Stars*, 5, CE, WO, HC, FA, MB; *Chains*, 0; *Triangles*, 0; *Inter-sexual Attraction*, 22.

EVOLUTION OF GROUPS

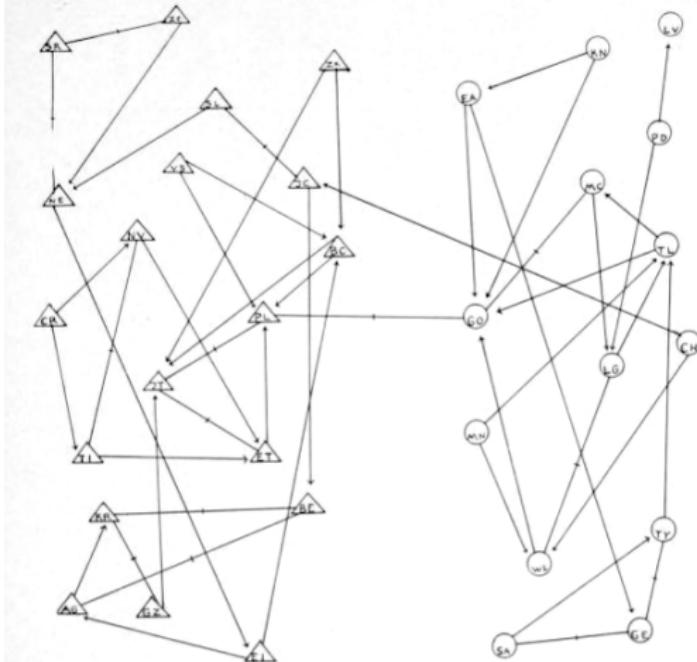


CLASS STRUCTURE, 2ND GRADE

14 boys and 14 girls. *Unchosen*, 9, WI, KP, MG, AT, FS, CN, CR, MR, SH; *Pairs*, 11, ZV-MK, MK-LN, OW-ZI, GR-LL, ZI-JM, HN-CM, SL-JN, JN-PO, PO-SL, HF-BE, GL-GU; *Stars*, 2, SL, PO; *Chains*, 0; *Triangles*, 1, SL-JN-PO; *Inter-sexual Attractions*, 5.

Moreno, "Who Shall Survive?" (1934)

EVOLUTION OF GROUPS

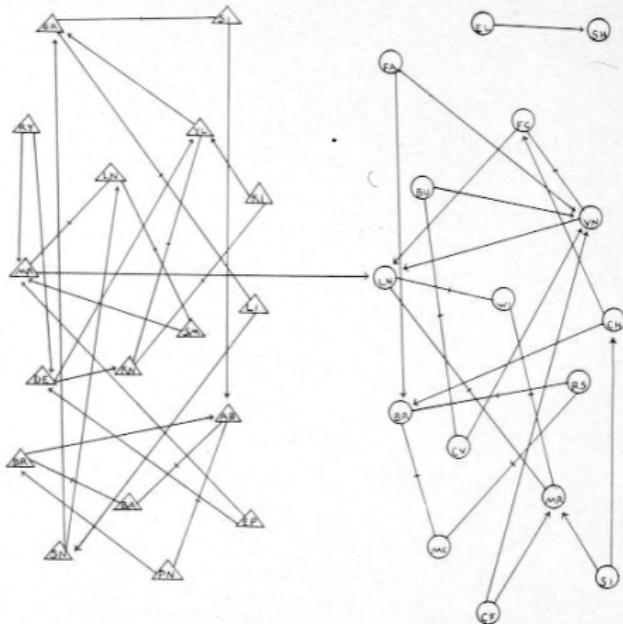


CLASS STRUCTURE, 3RD GRADE

19 boys and 14 girls. *Unchosen*, 7, VS, CR, CH, MN, PO, KN, ZK;
Pairs, 14, SR-ZC, SR-NE, SL-JC, NV-TI, PL-JT, JT-ET, KR-BE,
 BE-AG, RR-GZ, PL-GO, GO-MC, WL-LG, SA-GE, GE-TY; *Stars*, 3,
 GO, PL, JT; *Chains*, 1, ET-JT-PL-GO-MC; *Triangles*, 0; *Inter-sexual
 Attraction*s, 3.

Moreno, "Who Shall Survive?" (1934)

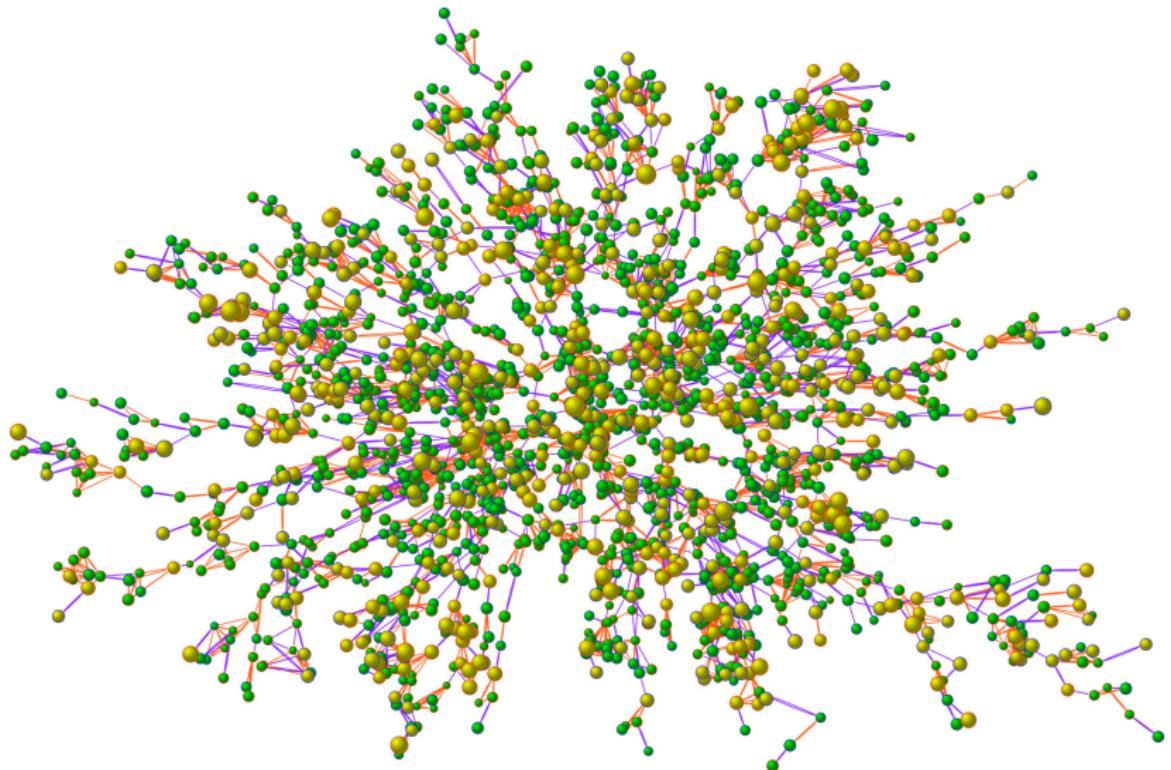
EVOLUTION OF GROUPS



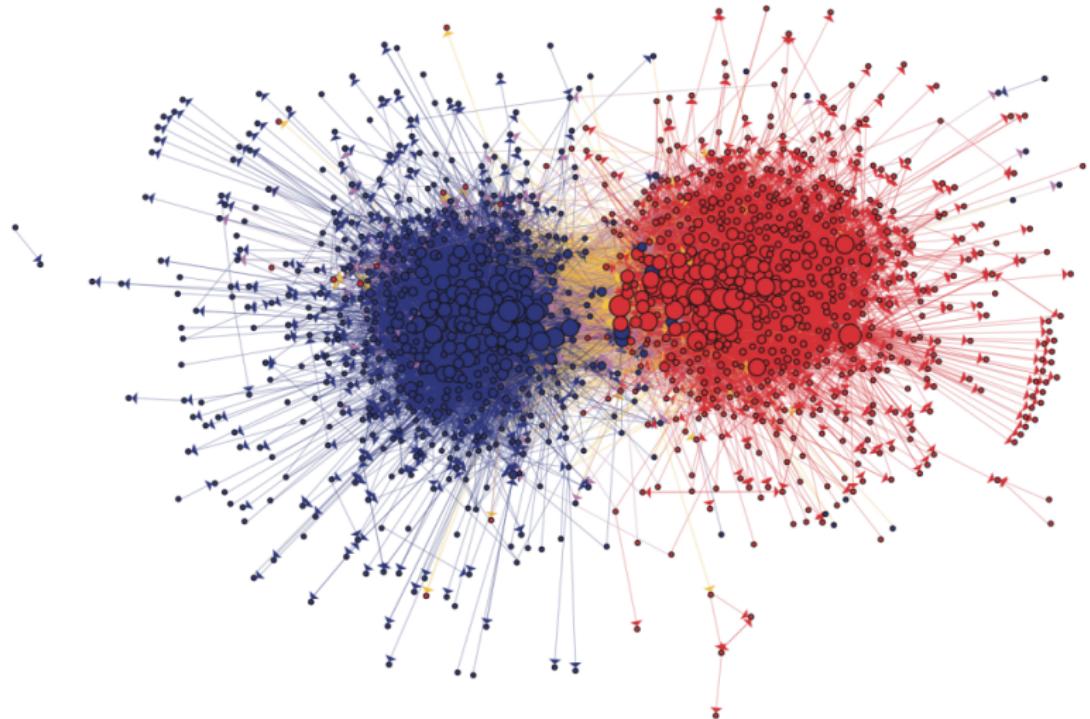
CLASS STRUCTURE, 4TH GRADE

17 boys and 16 girls. *Unchosen*, 6, EP, RY, EL, FA, SI, CF; *Pairs*, 17, GR-SI, GR-LI, MR-LN, LN-SM, YL-KN, AB-BA, BA-BR, KI-KN, AB-PN, FC-VN, BU-CV, LN-WI, LN-MR, BR-MC, BR-RS, WI-MR, MC-RS; *Stars*, 2, LN, VN; *Chains*, 0; *Triangles*, 2, BR-RS-MC; LN-WI-MR; *Inter-sexual Attractions*, 1.

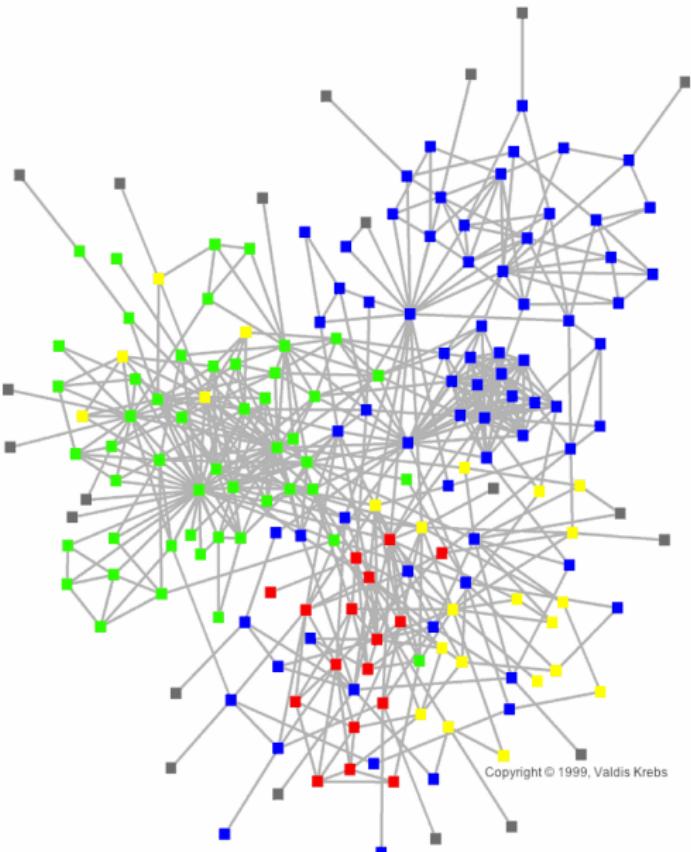
Moreno, "Who Shall Survive?" (1934)



Christakis & Fowler, NEJM, 2007

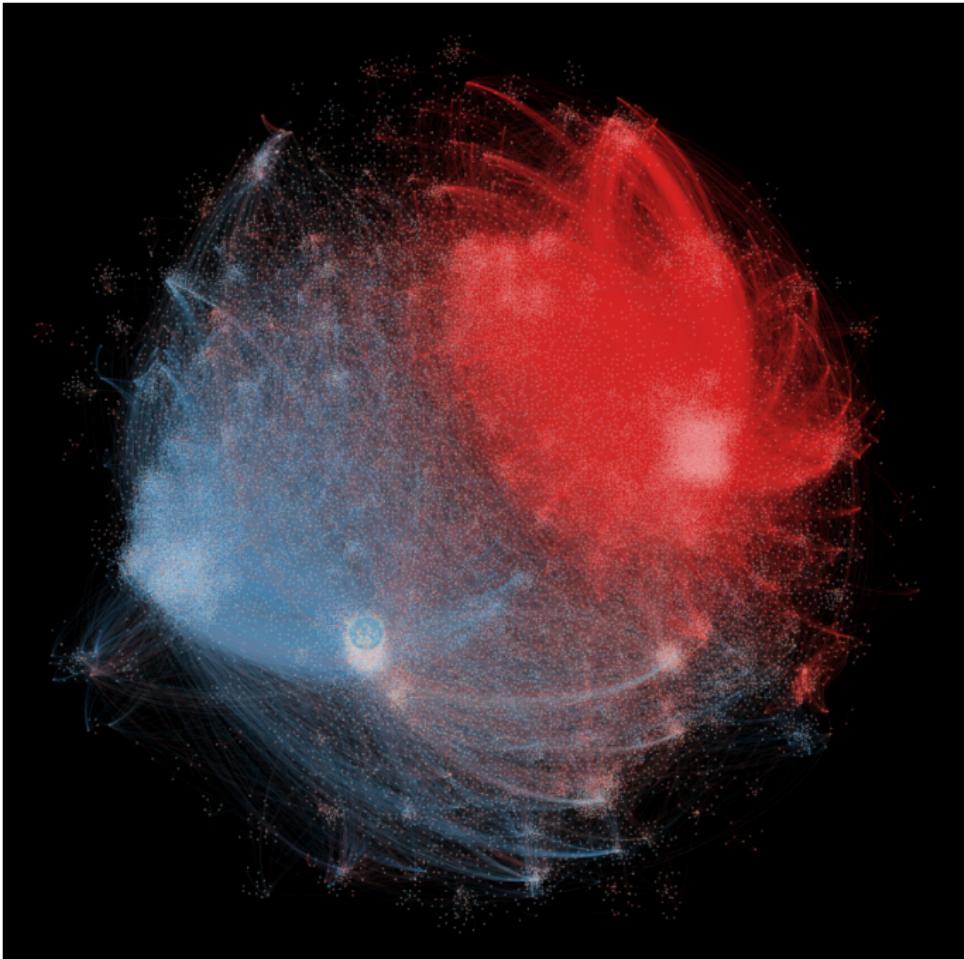


Adamic & Glance, 2004, IWLD

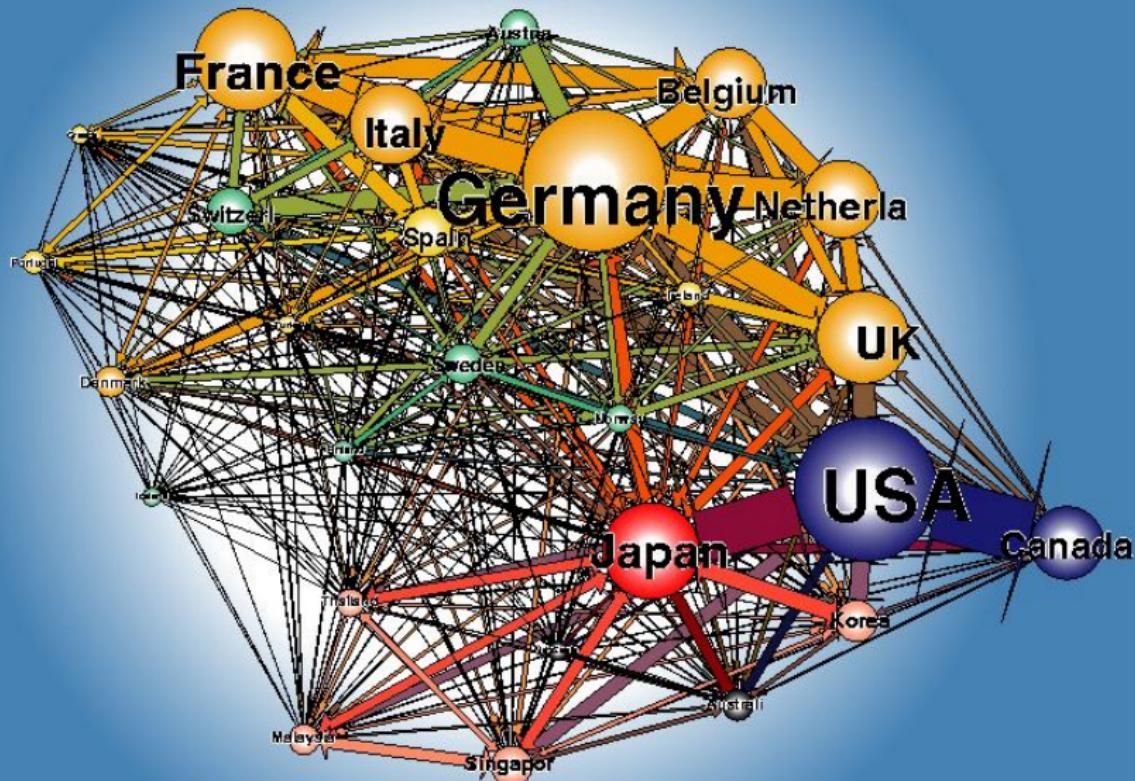


Copyright © 1999, Valdis Krebs

Email network of a company



Barbera et al, 2015, Psychological Science



Social Network Analysis

Today:

- ▶ Definitions
- ▶ Importing network data into R
- ▶ Network visualization
- ▶ Node and network-level summary statistics

Big Data course next week

- ▶ Community detection in large-scale networks
- ▶ Network decomposition techniques
- ▶ Modeling network properties: Exponential random graph models
- ▶ Large-scale network visualization with Gephi

Basic concepts

- ▶ **Node** (vertex): each of the units in the network
- ▶ **Edge** (tie): connection between nodes
 - ▶ Undirected: symmetric connection, represented by lines
 - ▶ Directed: imply direction, represented by arrows
- ▶ A **network** consists of a set of nodes and edges

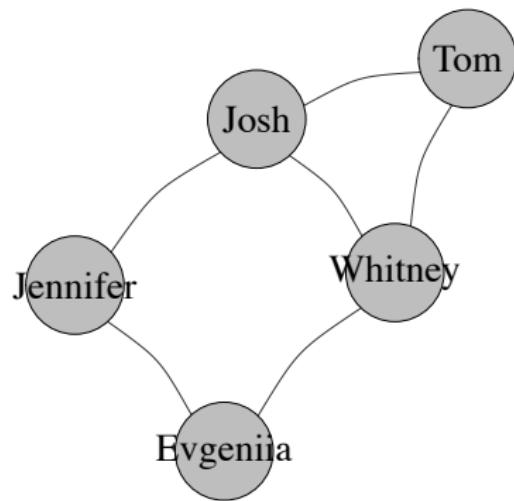
Basic concepts

A few examples:

- ▶ Classroom: students / friendships
- ▶ Twitter: users / retweets
- ▶ Academic literature: papers / citations
- ▶ Internet: websites / hyperlinks
- ▶ Trade: countries / trade flows
- ▶ Biology: neurons / connections

Basic concepts

Network Visualization

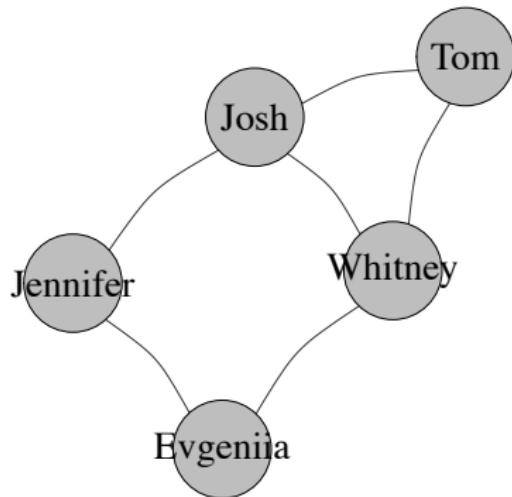


Adjacency Matrix

	J	J	E	W	T
J	0	1	1	0	0
J	1	0	0	1	1
E	1	0	0	1	0
W	0	1	1	0	1
T	0	1	0	1	0

Basic concepts

Network Visualization



Edgelist

	Node1	Node2
1	Jennifer	Josh
2	Jennifer	Evgeniia
3	Josh	Whitney
4	Josh	Tom
5	Whitney	Tom
6	Evgeniia	Whitney