

# Network mining: Historic data analysis and online social debates

Aristides Gionis

Department of Computer Science Aalto University

HELDIG Digital Humanities Summit 2017 Oct 18, 2017

## data mining group in Aalto

- algorithmic data analysis
- theory and applications
- topics of interest
  - pattern discovery
  - text mining
  - network analysis
  - smart cities

. . .



### research on family trees

- research problem: reconstruct family trees by linking vital records
- dataset: 10 million transcribed vital/parish records from Finland
  - births, marriages, burials, ...
  - mid-16th to early 19th cent.
- task: link birth records to parents' birth records
- challenges:
  - duplicate names
  - spelling variations
  - missing records

No _///_	Födde och Ødpte 1	Alisten Fitsanin	tar 18 <u>17</u>			
- 11	-	Million, Som at 1 Willia Am	1			
-	the any	Gardaport tom jelan	larent.			
14 11	2 Solt They	to Superman giles gite	Conser. reprisery			
	len I	and the patter factor	- Aging Haras	4 Section Bree	w find stille	to fites pile
. 19 5	a Martin Mary	a Busher Judah	Station .	And bear Ju	and Beater a	No care parts
	II (ciffued )	parters from shifter i	S all Silma	Southand Sing	and there is a	hope fine a
				ild p	ledge to to be a	Sens plant for
			6 11 10 Solt	they say	tichelp In lafor	Balan to tim
	21 2 3 Saly reals	liter how filst		Consigne the ly	and to Hidrag at	Conference allates
	1.1.1	Send Filian	1 1 1 m har link	Horton to	41 filon filons	1 I Tolan
		2 public find a	Jan Bay Sta			
	2. 3. W. Holy	ture 1 State long	The Road			
	7 3 23 5. 6.1	Child States of 19	4 the signal	ž		
	1100	tick at life trife, +	getting	1		
	de la com	- ind in	de to			

#### research on family trees

- research problem: reconstruct family trees by linking vital records
- computational aspects:
  - link records jointly
  - incorporate experts' feedback
- question: what can we learn from the family tree?
  - e.g., social stratification

#### relevant work:

- E. Malmi, S.Chawla, A. Gionis, "Lagrangian relaxations for multiple network alignment", DMKD 2017
- E. Malmi, A. Gionis, E. Terzi, "Active network alignment: A matching-based approach", CIKM 2017
- E. Malmi, M. Rasa, A. Gionis, "AncestryAI: a tool for exploring computationally inferred family trees", WWW Companion, 2017



### demo: ancestryai.cs.hut.fi



#### social media



- people use social media to
  - share information, express opinion, comment, interact, discuss, get personalized news feed

#### advantages

- no information barriers
- citizen journalism
- social connectivity
- democratization

• . . .

#### disadvantages

- harassment
- fake news
- echo chambers
- polarization
- . . .

#### research questions

- can we identify polarized discussions in social media?
- has polarization increased over time?
- how does collective attention impact polarization?
- can we design algorithms to help reduce polarization?



#### online societal debates - relevant work

- K. Garimella, A. Gionis, N. Parotsidis, N. Tatti, "Balancing information exposure in social networks", NIPS 2017
- K. Garimella, G. De Francisci Morales, A. Gionis, M. Mathioudakis, "The effect of collective attention on controversial debates on social media", International ACM Web Science 2017
- K. Garimella, G. De Francisci Morales, A. Gionis, M. Mathioudakis, *"Reducing controversy by connecting opposing views"*, ACM WSDM 2017
- K. Garimella, G. De Francisci Morales, A. Gionis, M. Mathioudakis, "Quantifying controversy in social media", ACM WSDM 2016

## thank you

