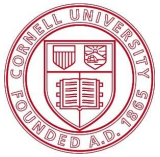
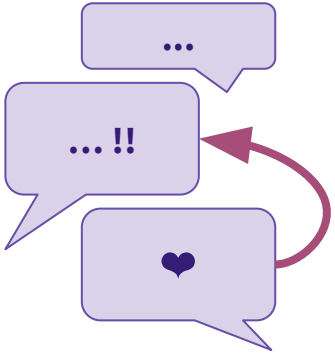
Characterizing Online Public Discussions Through Patterns of Participant Interactions

Justine Zhang, Cristian Danescu-Niculescu-Mizil,
Christina Sauper, Sean J. Taylor



Cornell University, Facebook

What does it mean to have a good discussion?

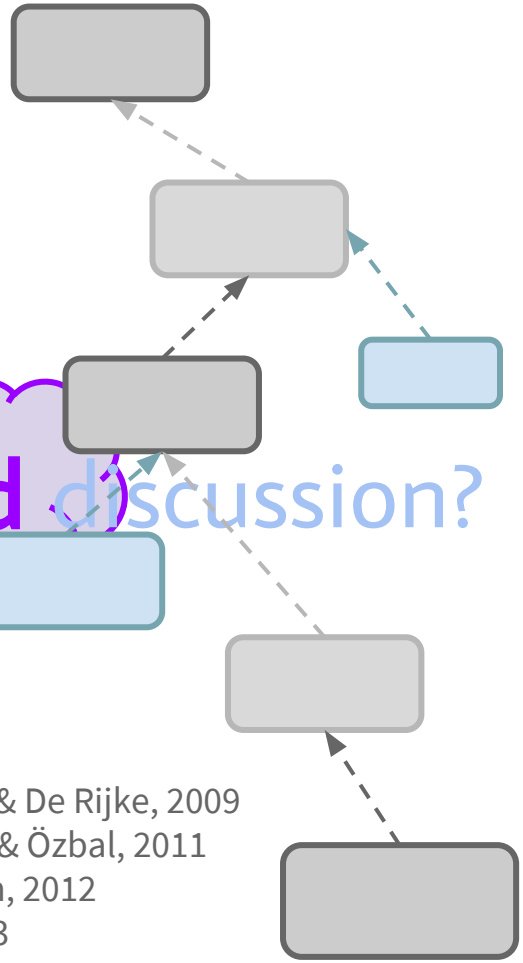


De Choudhury & Kıcıman, 2017
Yang & Kraut, 2017

Kittur et al., 2007
Cheng et al., 2016
Chandrasekharan et al., 2017
Wulczyn, Thain & Dixon, 2017
Kumar et al., 2018
Zhang et al., 2018

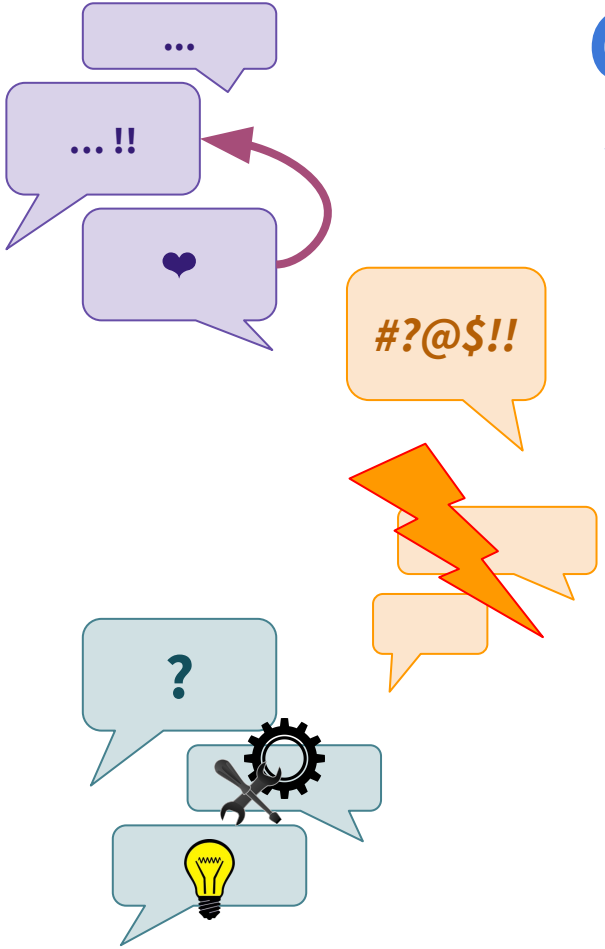
Gonzalez-Bailon, Kaltenbrunner & Banchs, 2010
Cranshaw & Kittur, 2011
Niculae & Danescu-Niculescu-Mizil, 2016

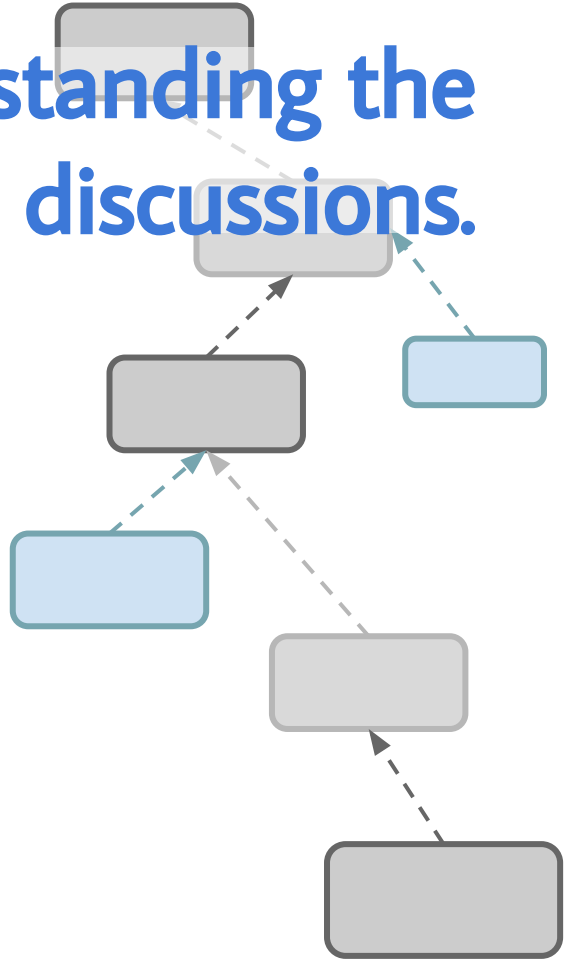
What does it mean to have a **good** discussion?



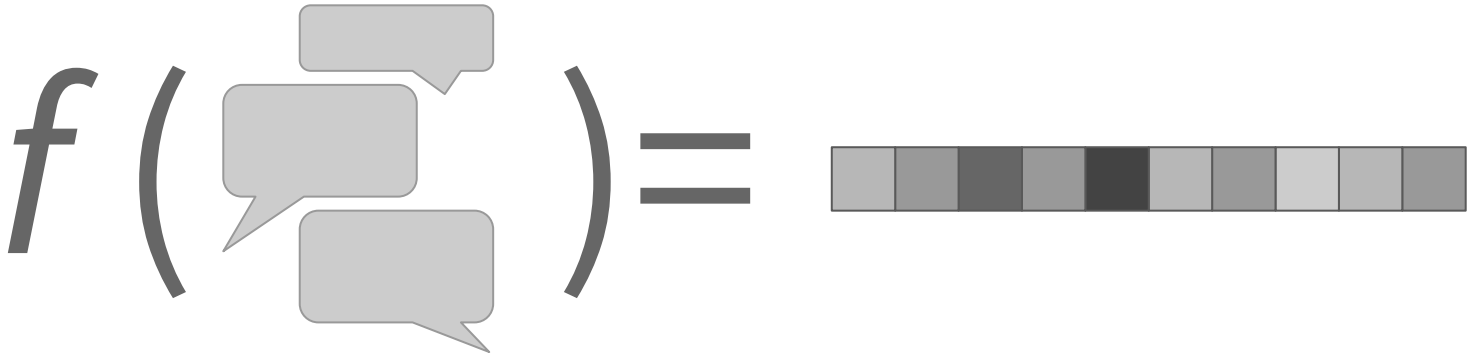
Tsagkias, Weerkamp & De Rijke, 2009
Guerini, Strapparava & Özbal, 2011
Arzti, Pantel & Gamon, 2012
Backstrom et al., 2013

Our goal: understanding the variation across discussions.





Our goal: understanding the variation across discussions.



Our goal: understanding the variation across discussions.

929,041 discussion threads with 10+ comments

The New York Times
14 October at 07:25 ·

We Could All Use a Little Snail Mail Right Now

What the world needs now? Handwritten cards and letters.

L

L: I disagree, snail mail causes gas fumes from all the mail delivery trucks.

S

S: I hope you make all your clothing from hemp since you're so worried about pollution.

P

P: Doing something for the environment is better than doing nothing.

M

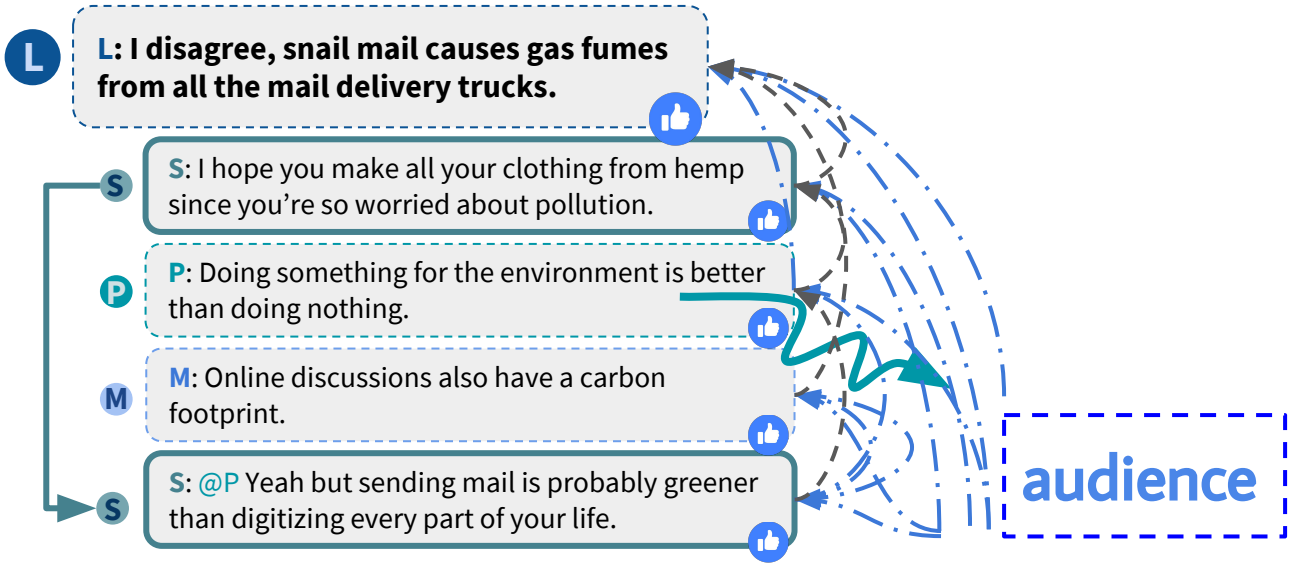
M: Online discussions also have a carbon footprint.

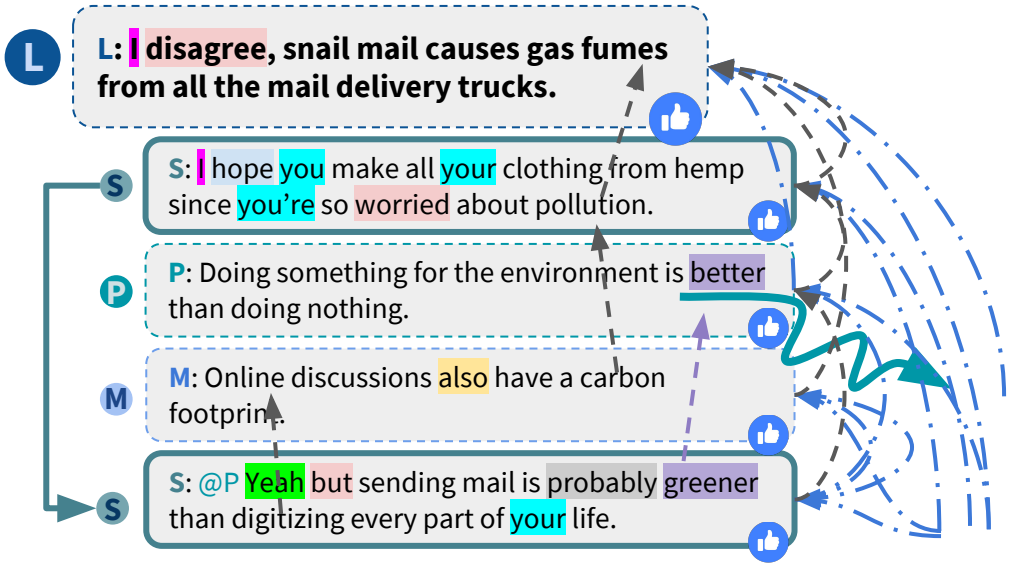
S

S: @P Yeah but sending mail is probably greener than digitizing every part of your life.

** actual Facebook discussion*

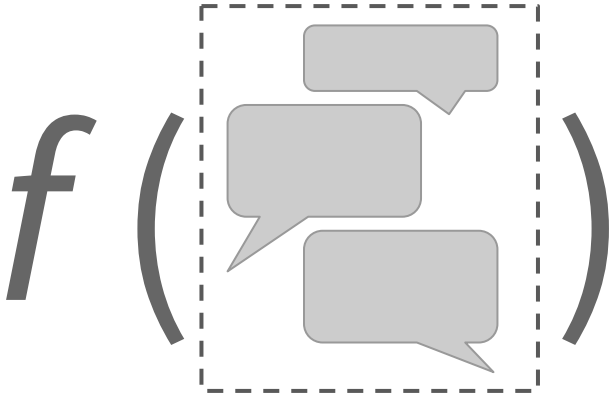
The New York Times
14 October at 07:25 ·
1,724
195
372
We Could All Use a Little Snail Mail Right Now
What the world needs now? Handwritten cards and letters.


L
L: I disagree, snail mail causes gas fumes from all the mail delivery trucks.
S
S: I hope you make all your clothing from hemp since you're so worried about pollution.
P
P: Doing something for the environment is better than doing nothing.
M
M: Online discussions also have a carbon footprint.
S
S: @P Yeah but sending mail is probably greener than digitizing every part of your life.
audience

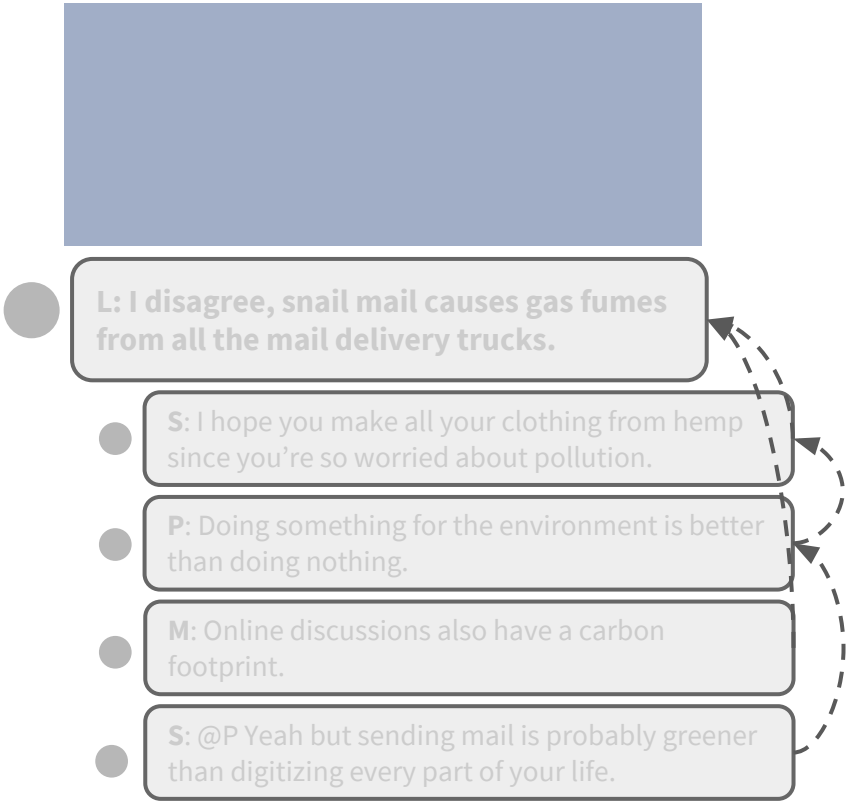


Ritter, Cherry & Dolan, 2010
West et al., 2014
Berry & Taylor, 2017
Zhang, Culbertson & Paritosh, 2017
Cheng, Fang & Ostendorf, 2017

Idea: **construct a representation of discussions.**



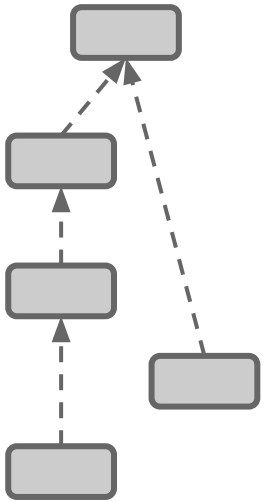
Representing discussions



Representing discussions

reply-tree view:

comments linked by **replies**

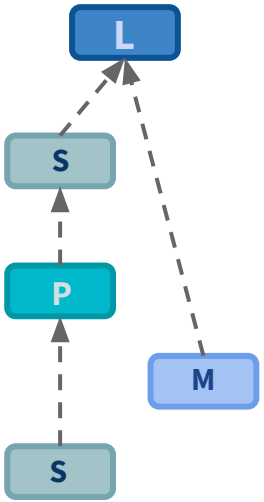


Gómez, Kaltenbrunner & López, 2008
Kumar, Mahdian & McGlohon, 2010
Laniado et al., 2011
Wang, Ye & Huberman, 2012
Nishi et al., 2016
Aragón et al., 2017

Representing discussions

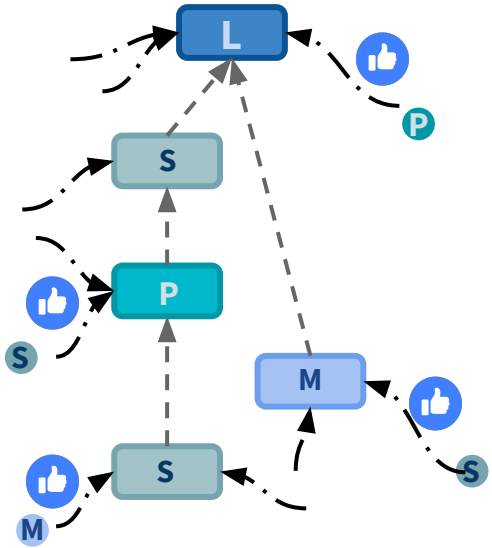
reply-tree view:

comments linked by **replies**



Representing discussions

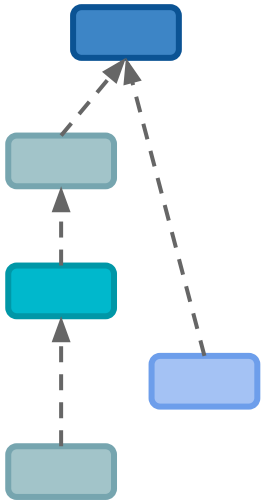
reply-tree view: ▭ - - - ►
comments linked by **replies**

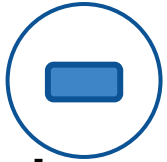


Representing discussions

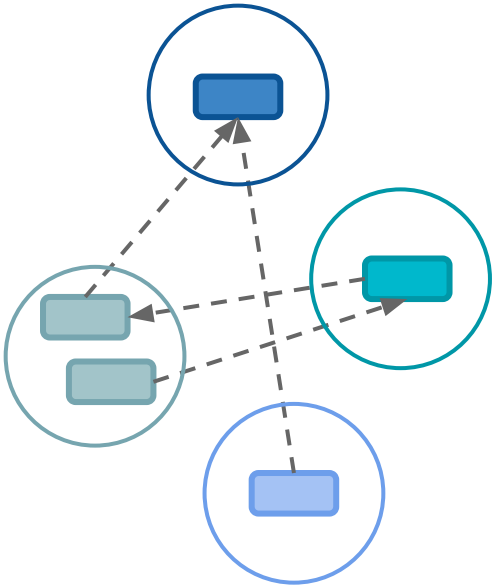
reply-tree view:

comments linked by **replies**



hypergraph view: 

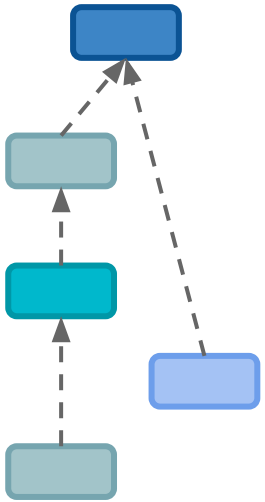
people linked by **relationships**

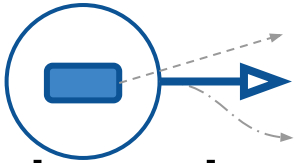


Representing discussions

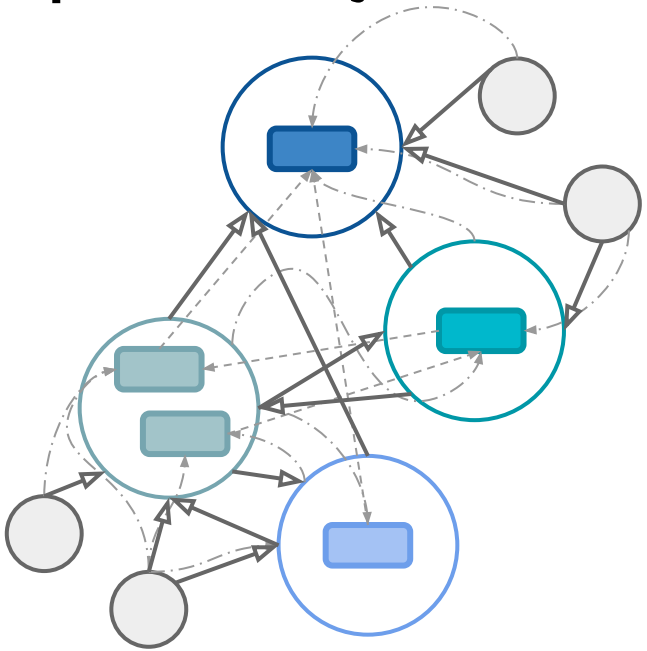
reply-tree view:

comments linked by **replies**

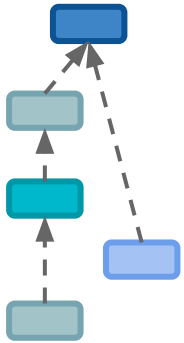


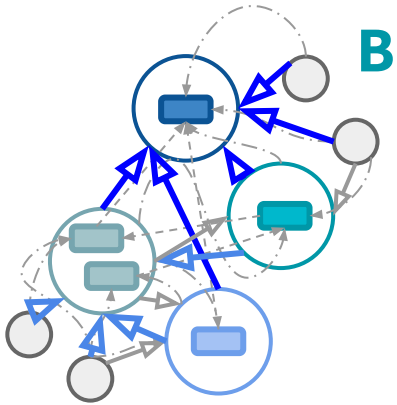
hypergraph view: 

people linked by **relationships**

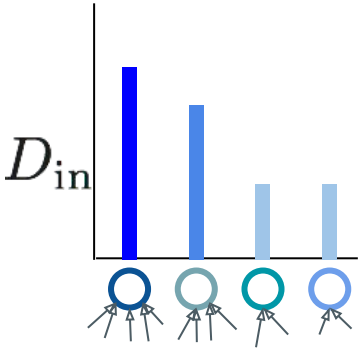


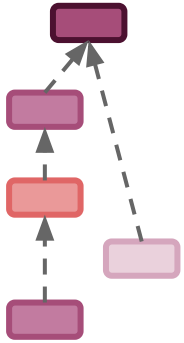
Describing discussions

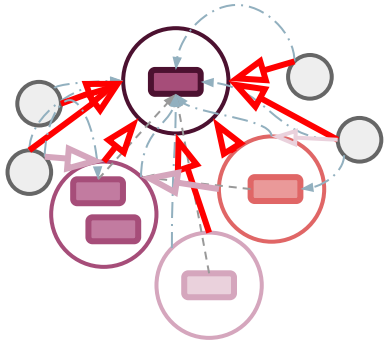




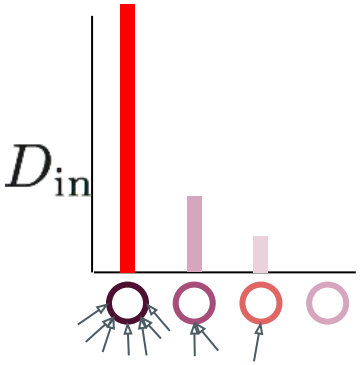
Balanced





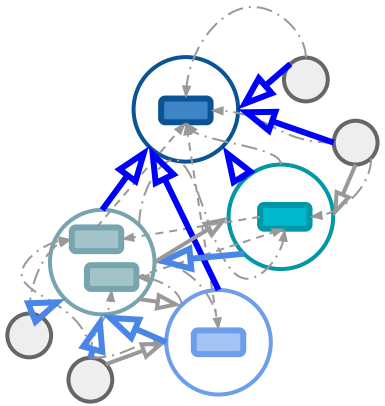


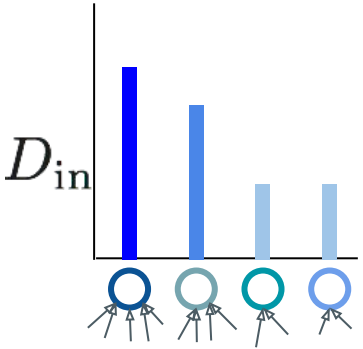
Skewed

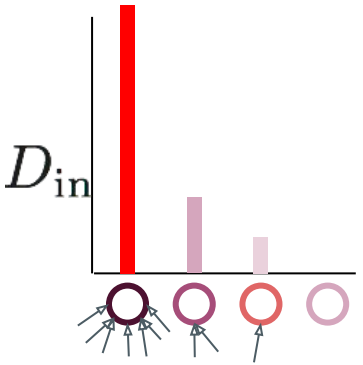


Describing discussions

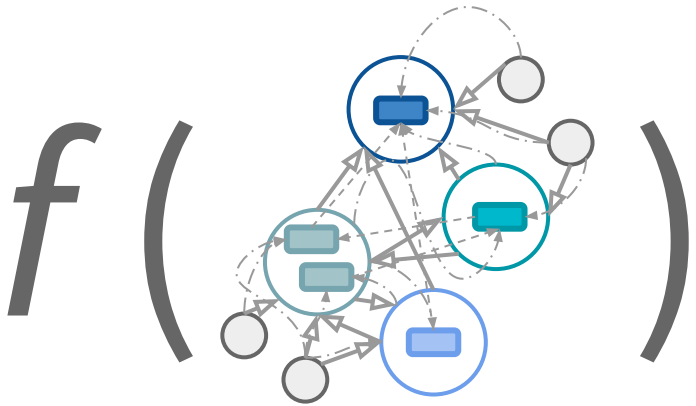
$$f\left(\quad\right) = \left[\frac{\max D_{\text{in}}}{\max_2 D_{\text{in}}}, \ldots\right]$$



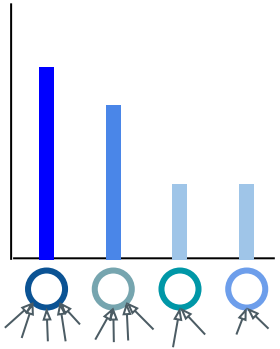




Describing discussions

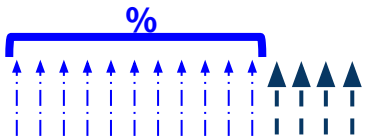


Degree distributions



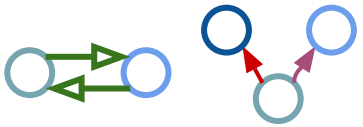
…

Edge distributions



…

Motif distributions

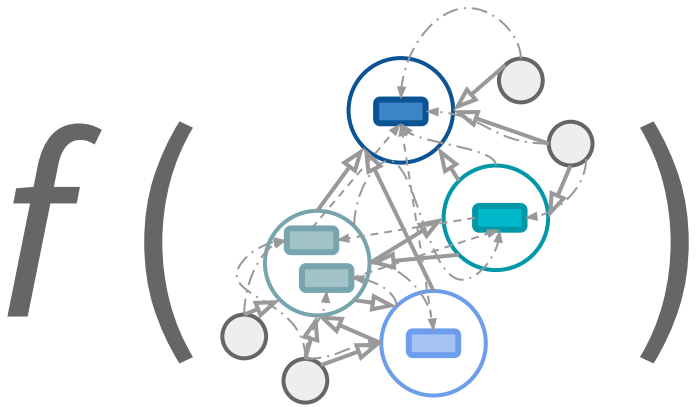


…

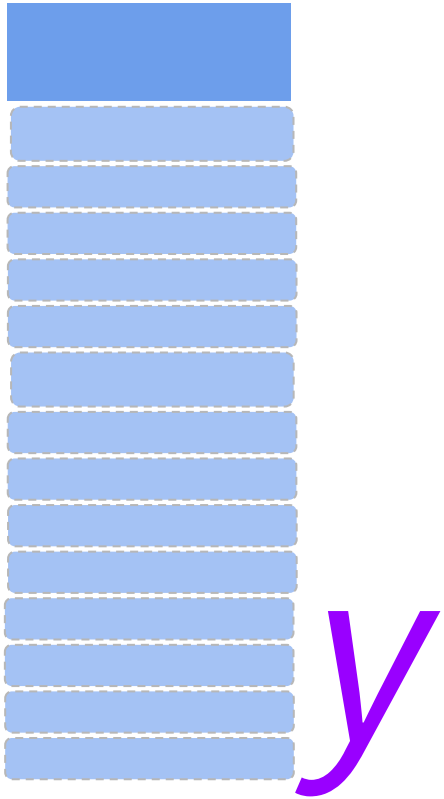
=

454 features

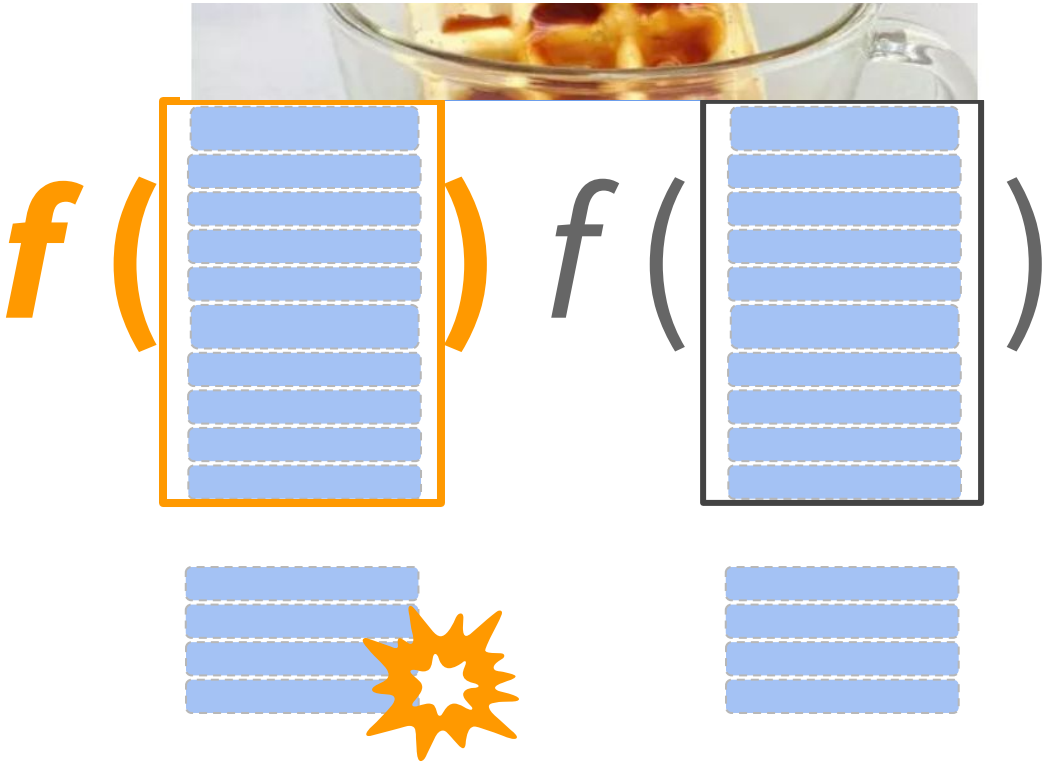
Application: predicting potential quality signals



=

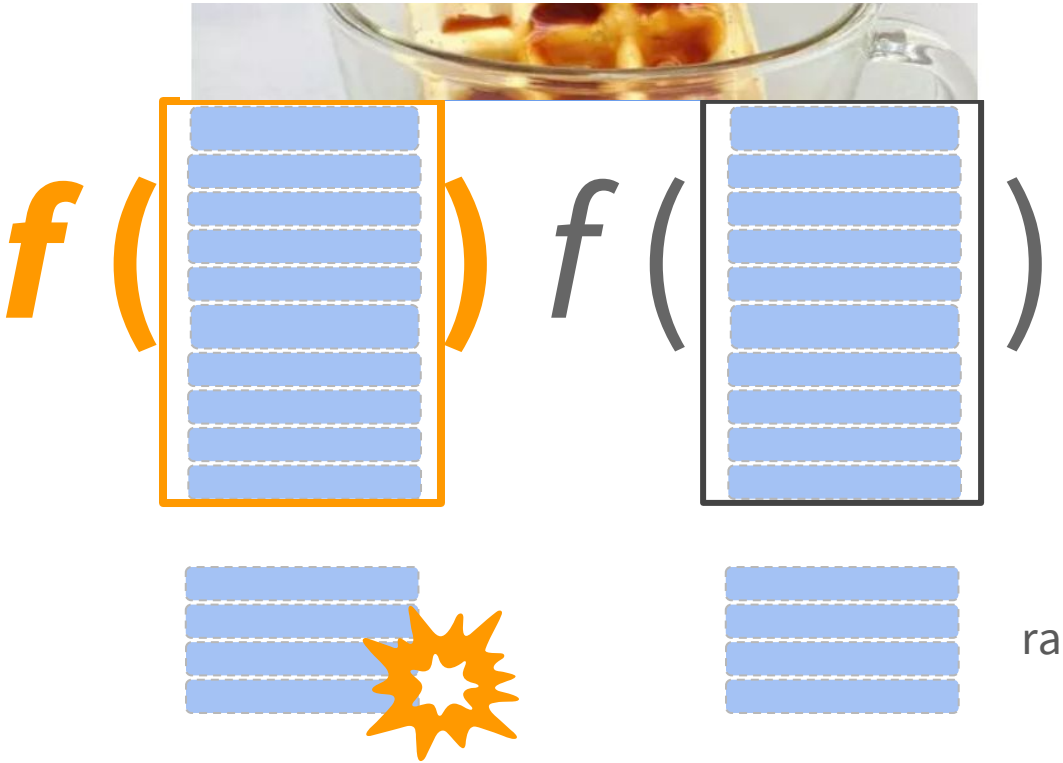


Application: predicting future antisocial behavior

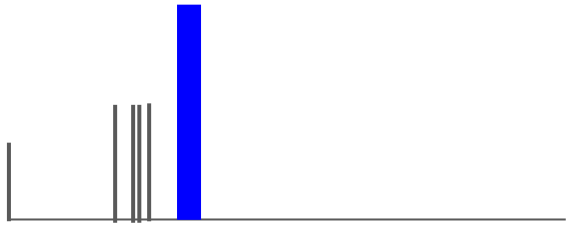


5,590 thread pairs

Application: predicting future antisocial behavior



hypergraph-derived features: 64.4%



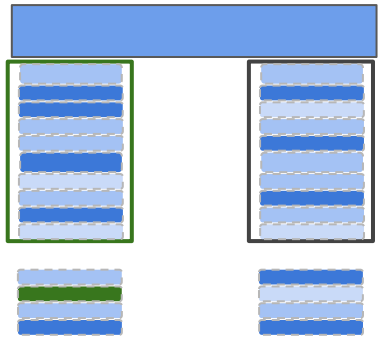
random: 50%

language/ reply-tree/ rate/views/ arrival patterns: ~60%

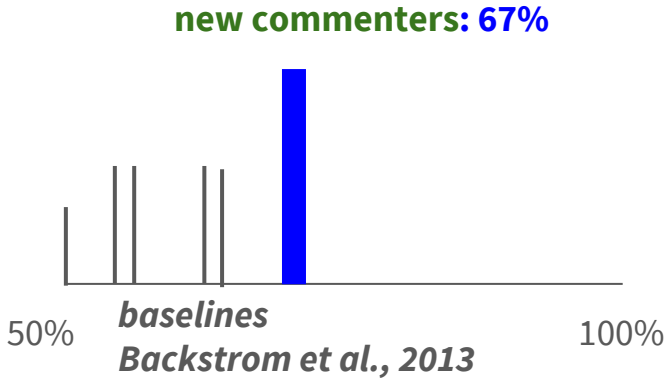
100%

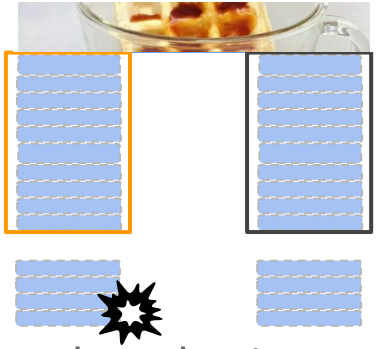
5,590 thread pairs

Application: predicting future antisocial behavior, etc.

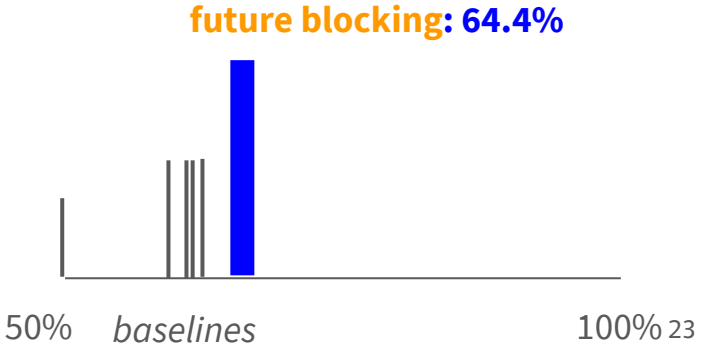


14,739 thread pairs

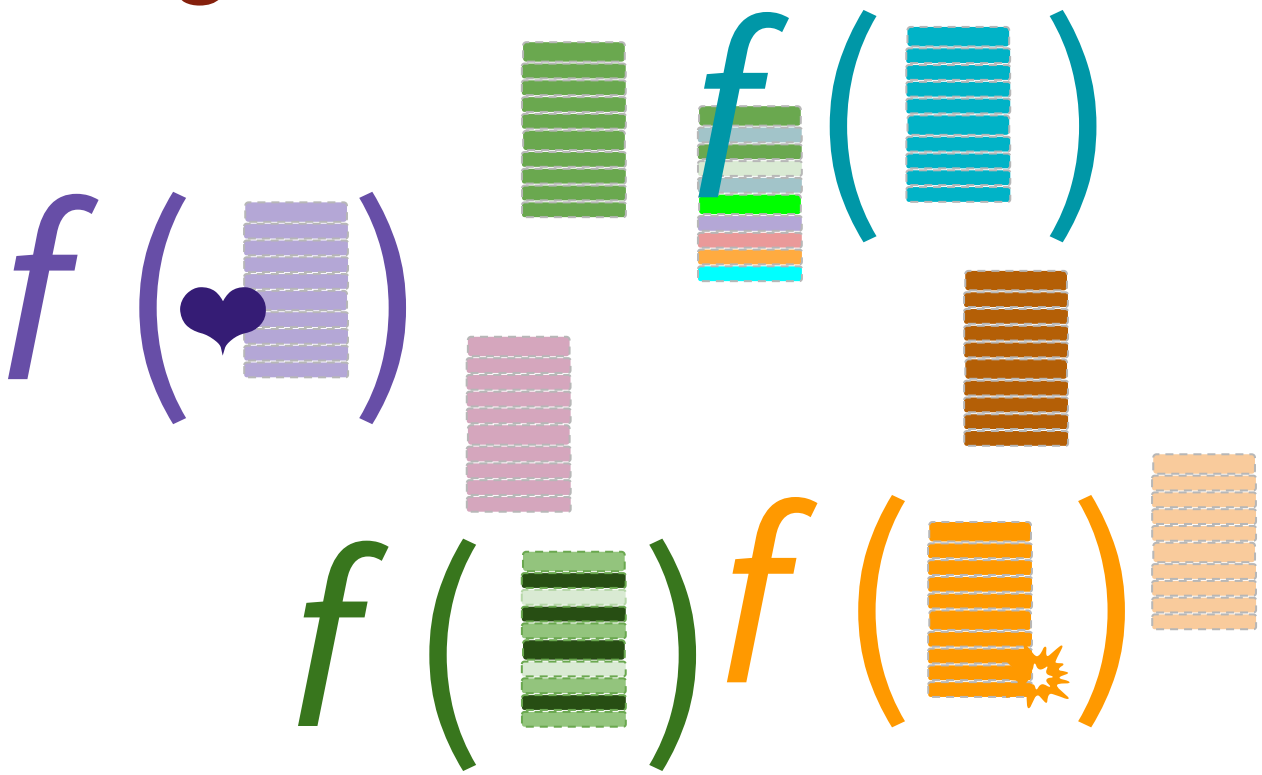




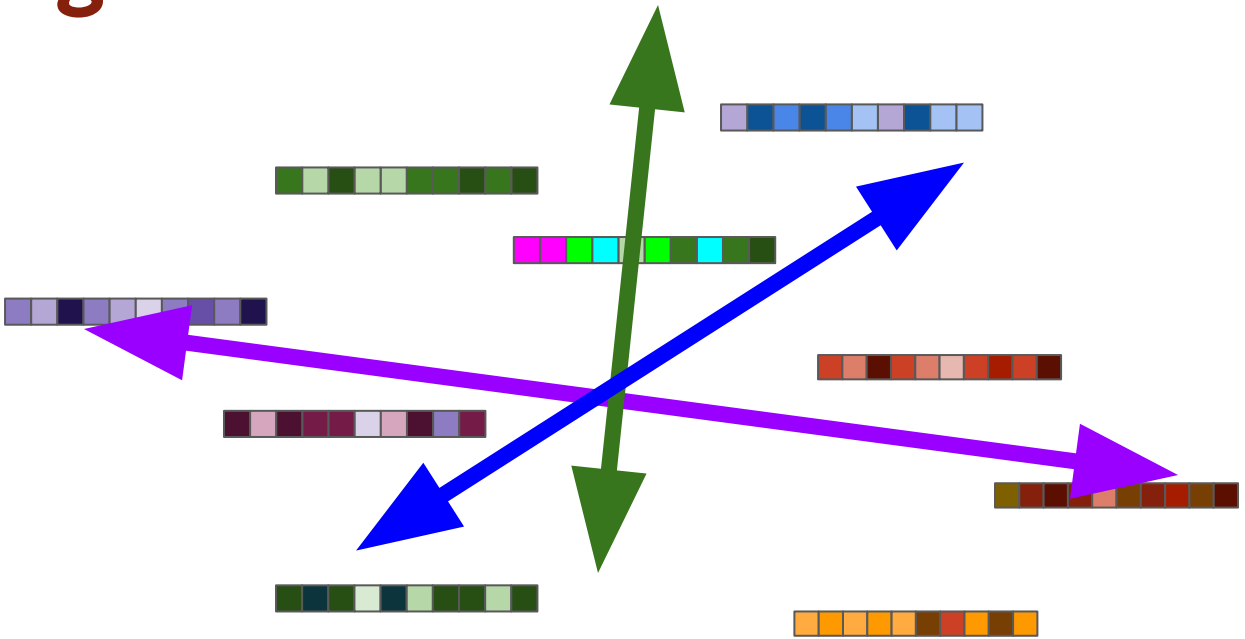
5,590 thread pairs



Inferring salient axes of variation

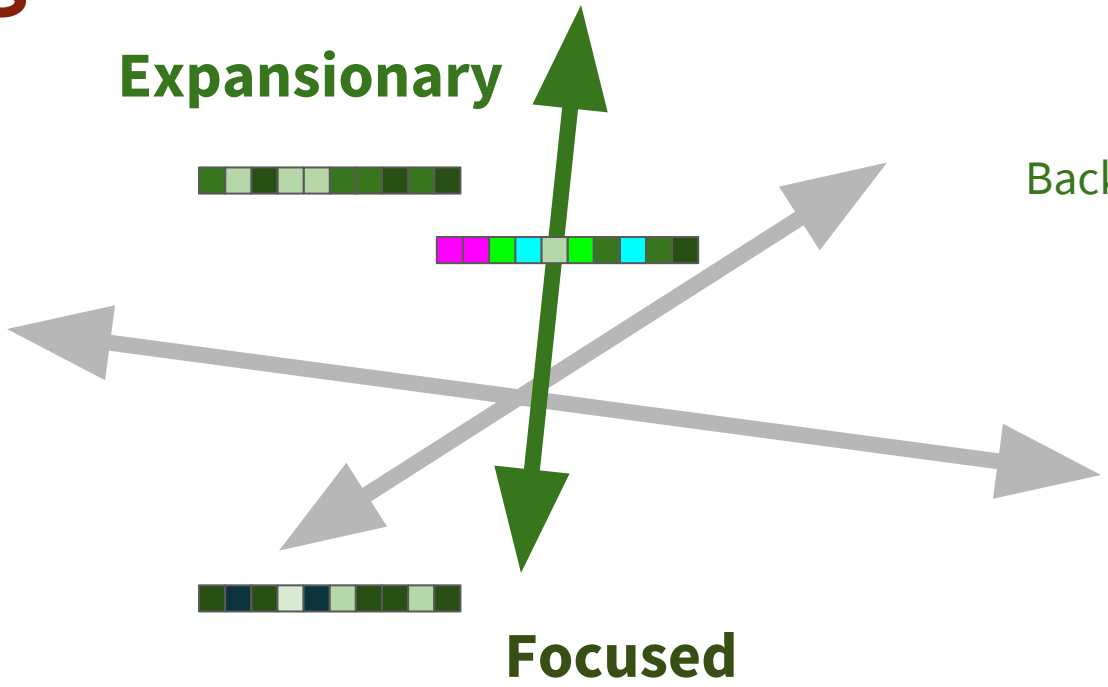


Inferring salient axes of variation



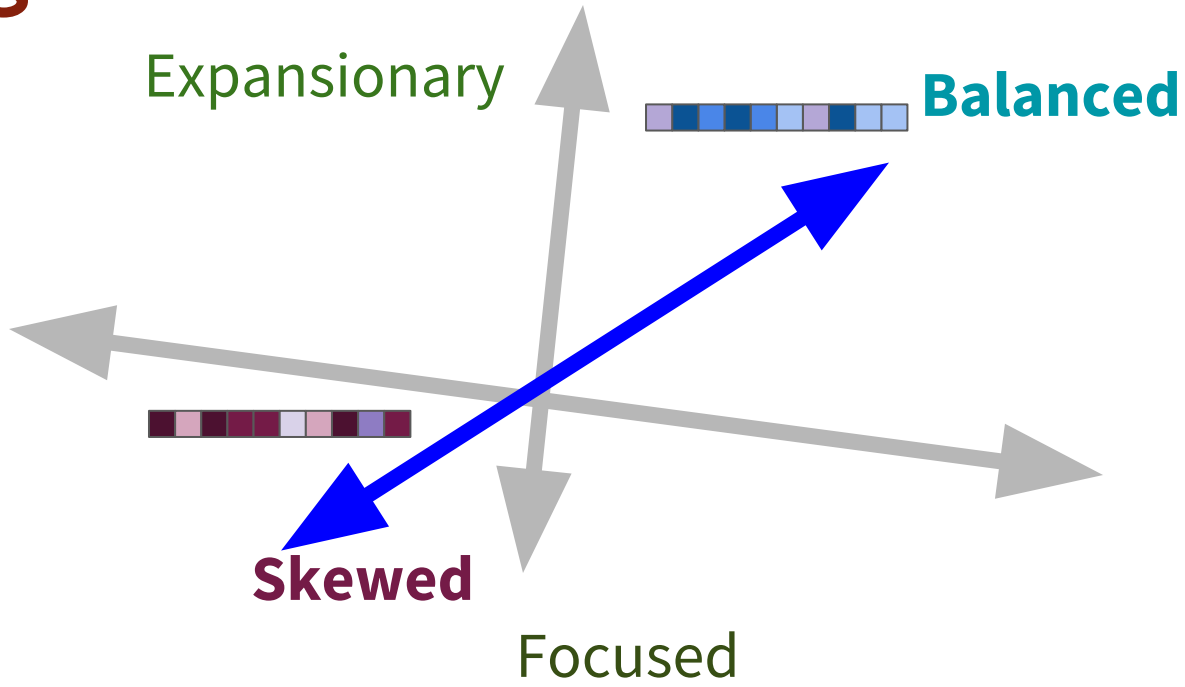
(using singular value decomposition)

Inferring salient axes of variation



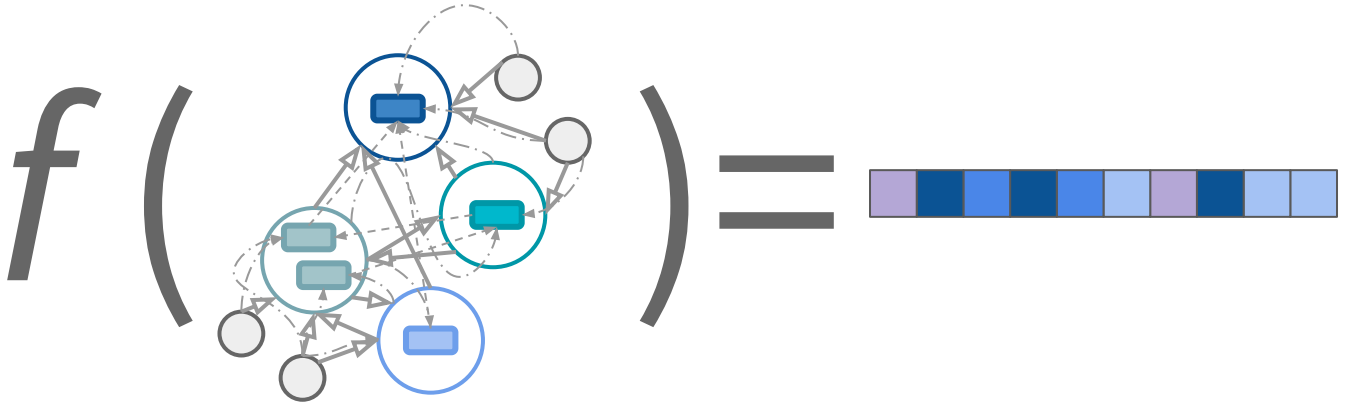
Backstrom et al., 2013

Inferring salient axes of variation

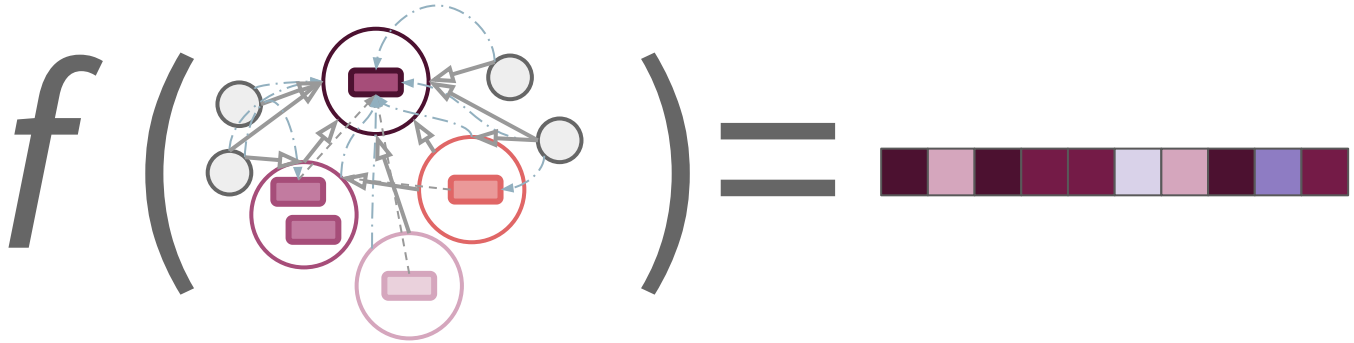


Inferring salient axes of variation

Balanced



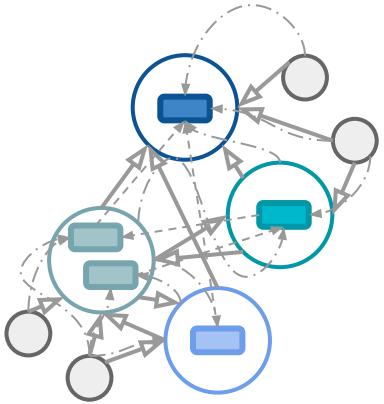
Skewed



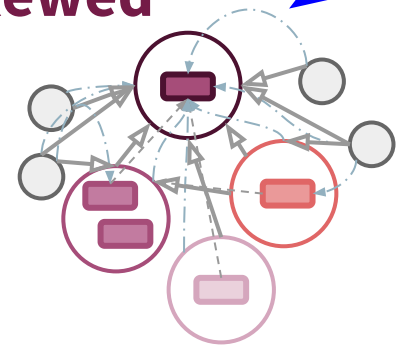
Inferring salient axes of variation

8,901 public Facebook Pages

Balanced



Skewed



Inferring salient axes of variation

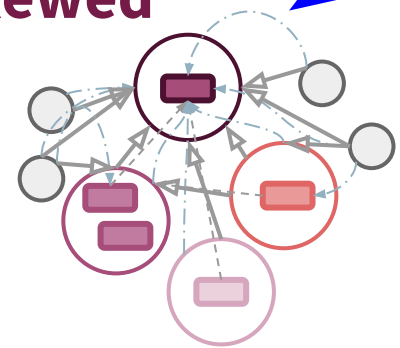
Sean Hannity

Breitbart

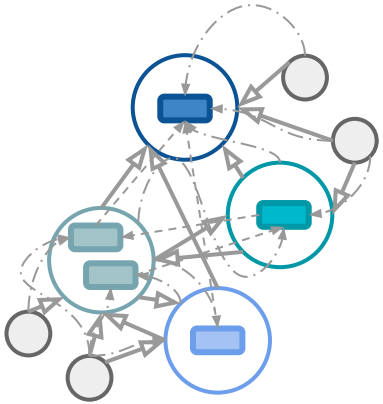
Robert Reich

FOX News [...]

Skewed



Balanced



ESPN

NFL

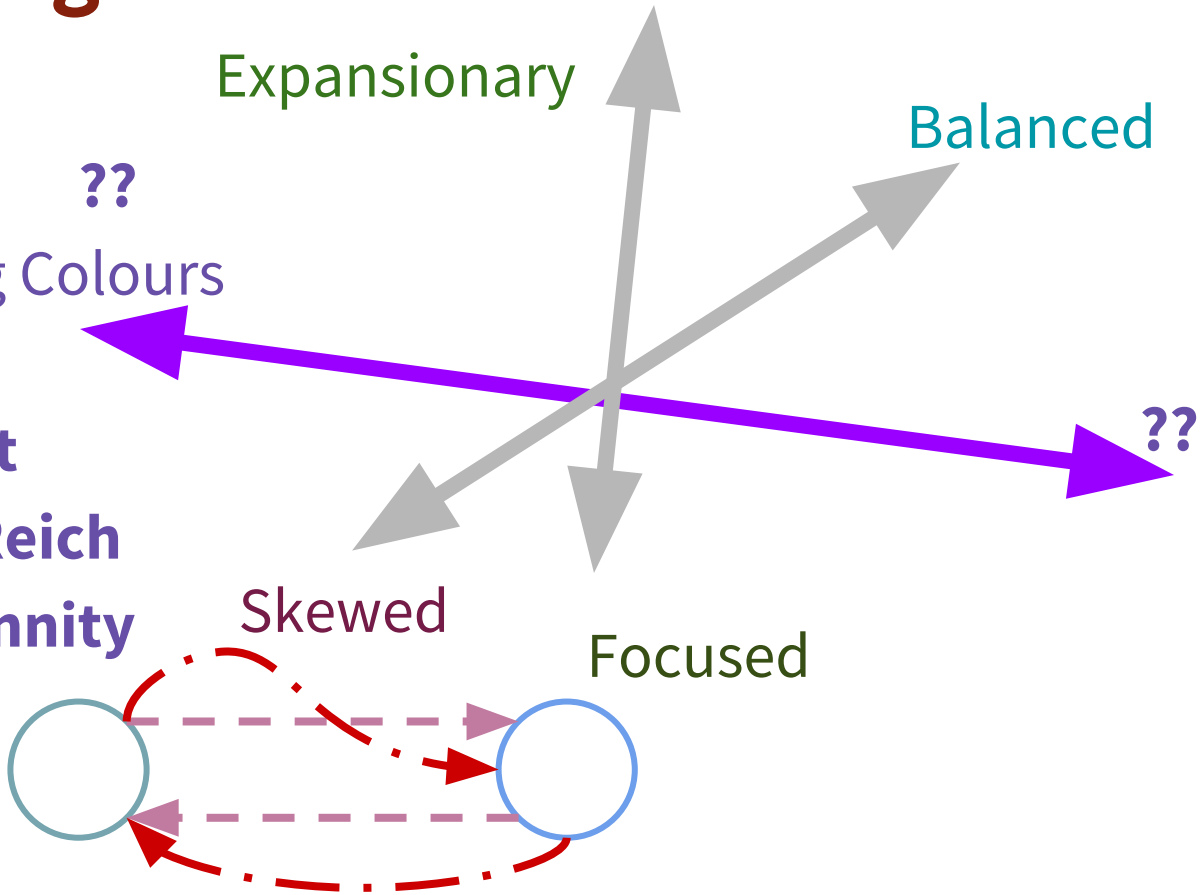
IGN

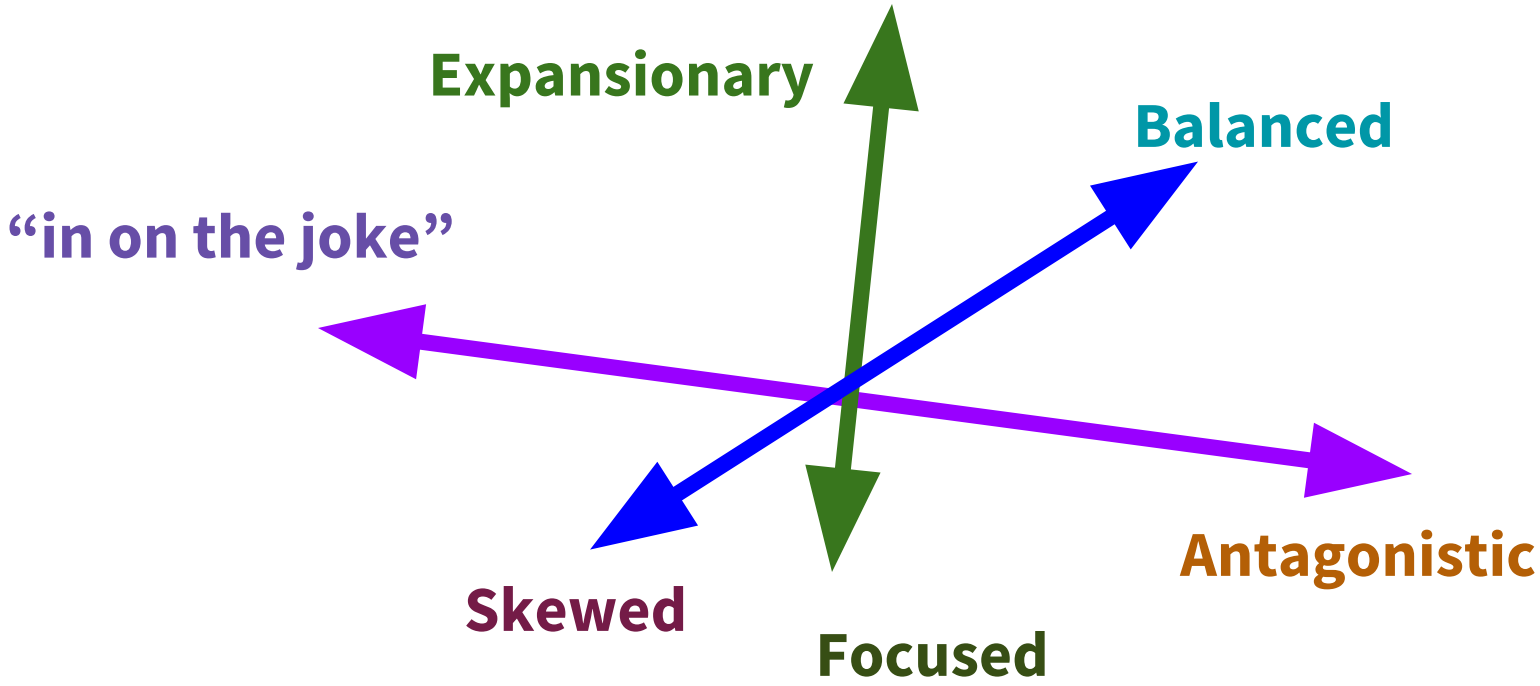
L'EQUIPE,

ΣΠΟΡ FM 94.6,

RÉCORD México [...]

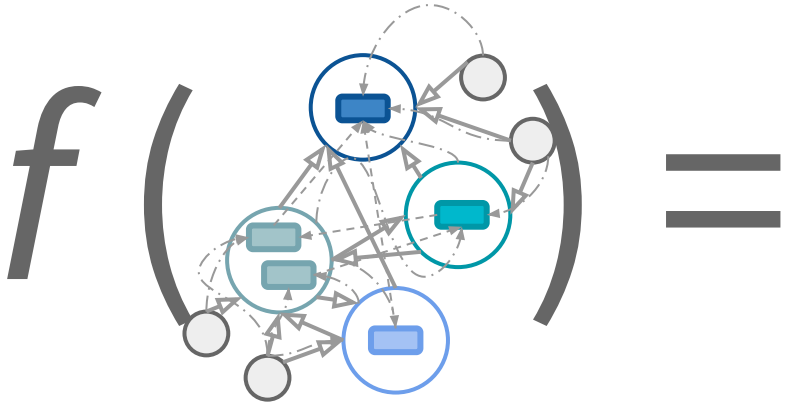
Inferring salient axes of variation

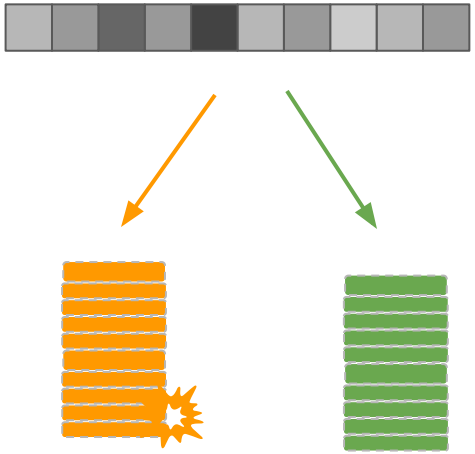




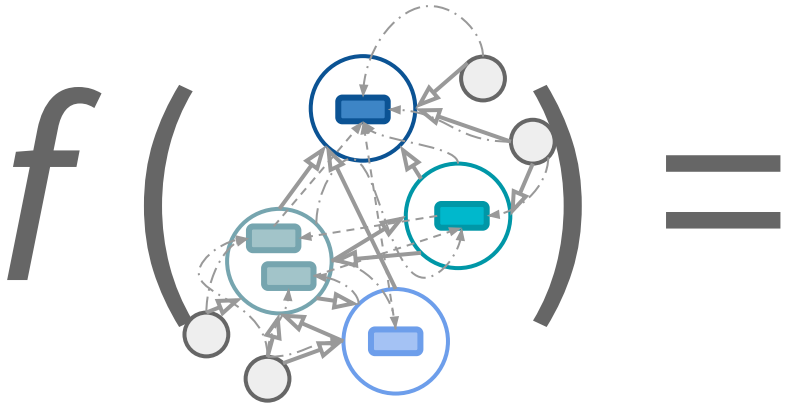
(see paper for other axes)

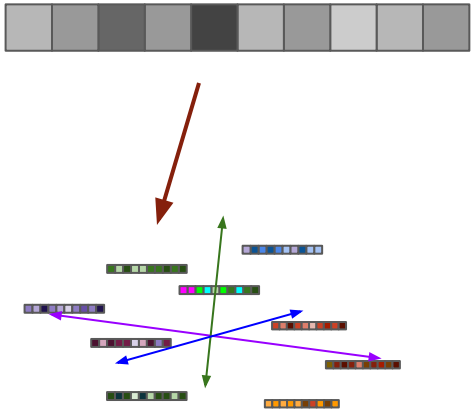
Conclusion



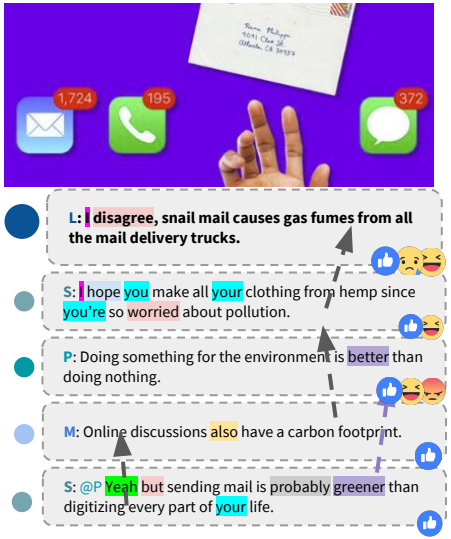


Conclusion





Future work: richer representations



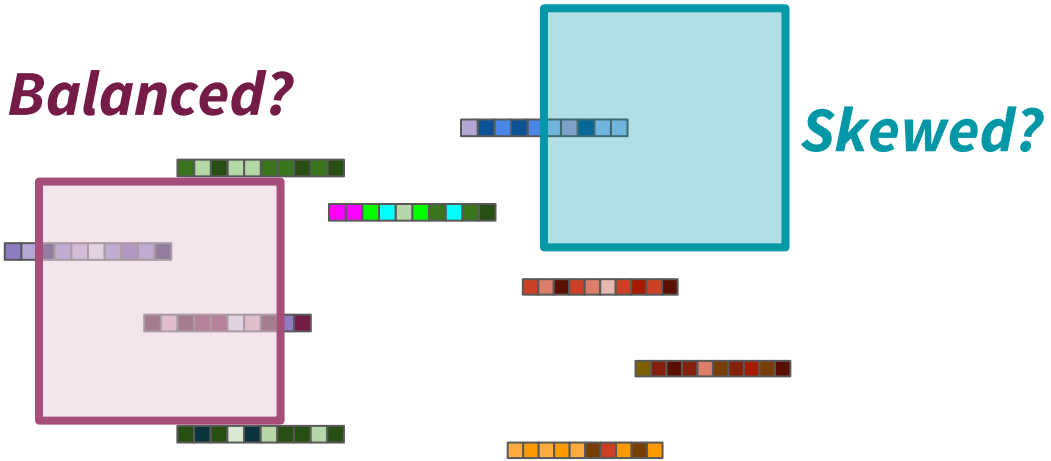
$$f\left(\quad\right) =$$

other actions
language
participant information

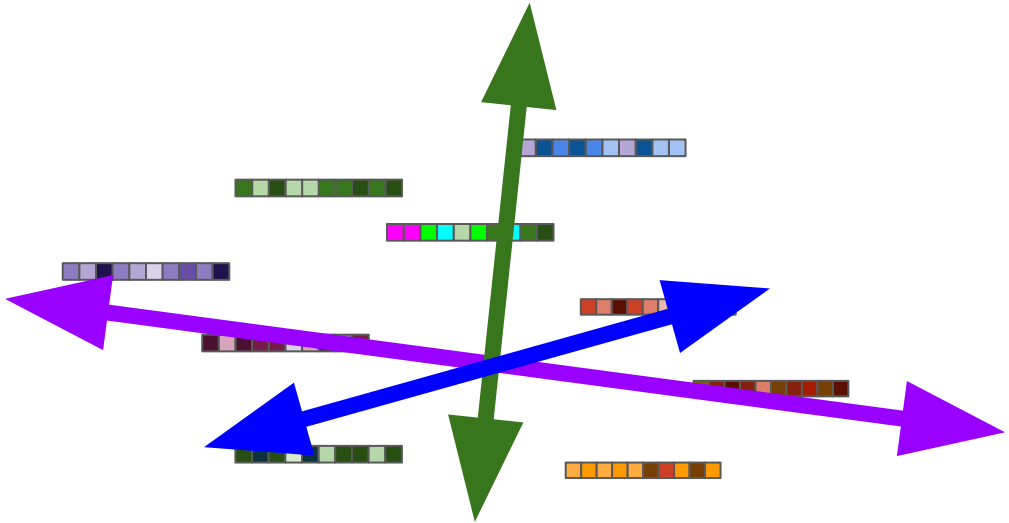
Future work: other platforms

Code:
Cornell Conversation Analysis Toolkit
convokit.cornell.edu

Future work:
understanding the impact of discussions



Future work:
understanding the impact of discussions



Code:
Cornell Conversation Analysis Toolkit
convokit.cornell.edu