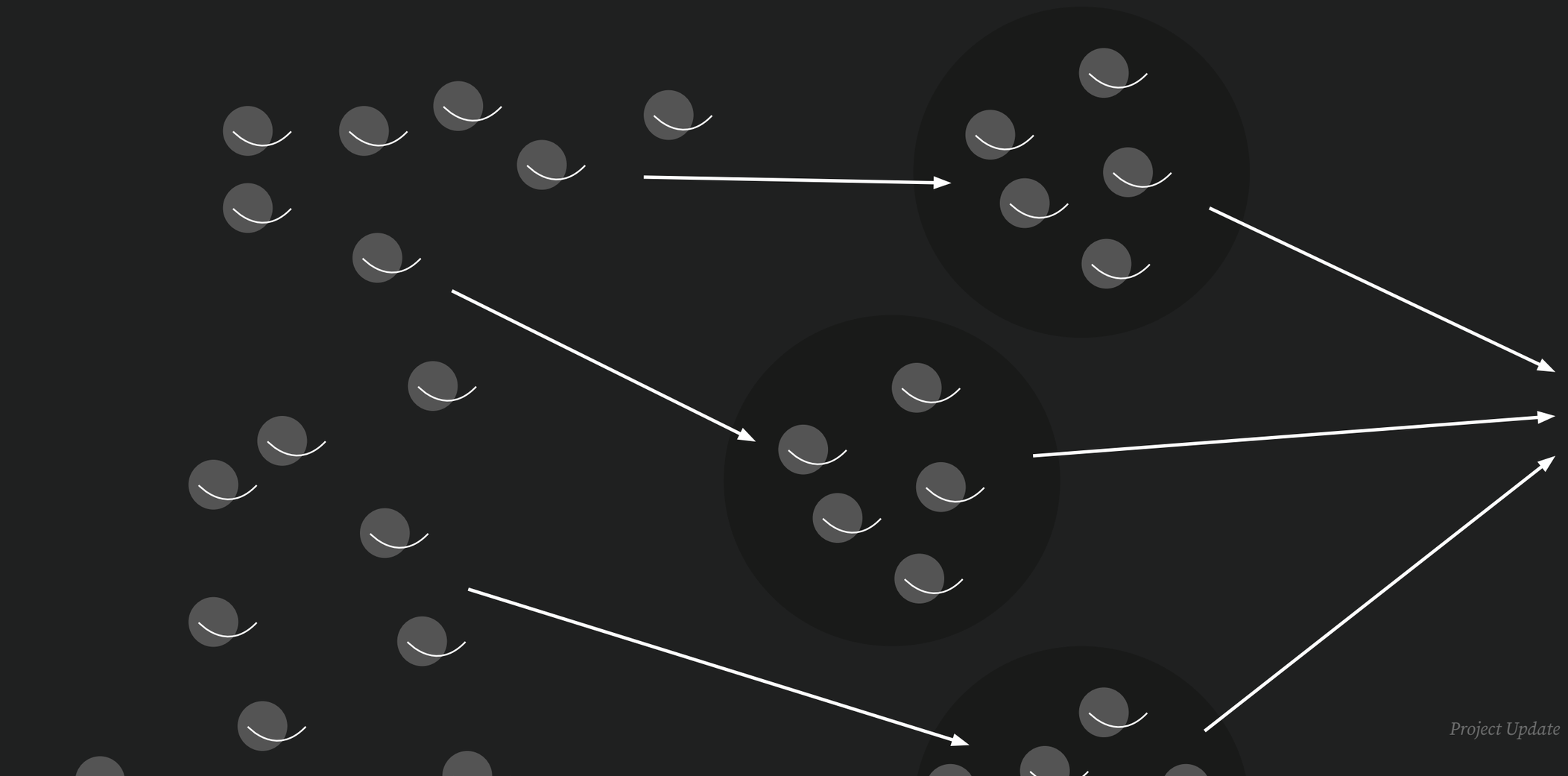
Facilitating Large Scale Discussions: Project Update

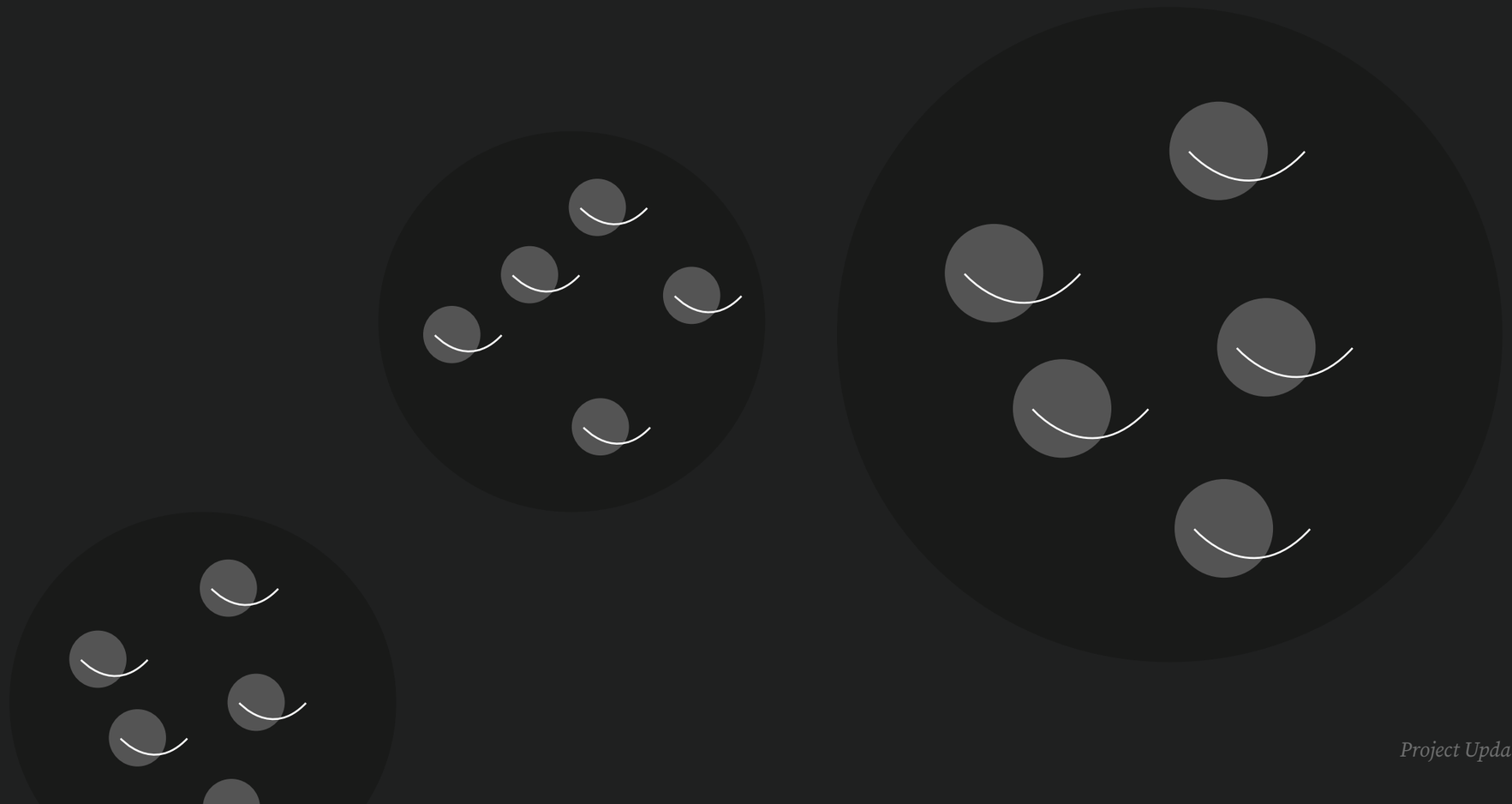
Christian Broms

November 18, 2020

Making large discussions manageable



Refined focus on discussion dynamics



What domain?

Selecting a domain might help if there were data available to compare with model

Designing partition methods is uncommon; studies not usually about how small groups are formed

No data sets (to my knowledge) that seem to fit

Better to keep it abstract, then try applying it to specific instances

What methodology?

Modeling how *desired* outcome changes influence *actual* outcome

Create a similarity metric that represents relatedness of desired outcomes: *homogeneity*

The agents

Have some desired outcome (perhaps *perspective* or *agenda*)

Every perspective is slightly different; there's nuance

Overarching goal might be similar, but individual variation in desired implementation

The agents' representation

The agents' representation

[0.55, 0.2, 0.1, 0.1, 0.05]

[0.70, 0.1, 0.05, 0.1, 0.05]

Each perspective is a normalized vector of features

The agents' similarity

A

B

Can determine relatedness using cosine similarity

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

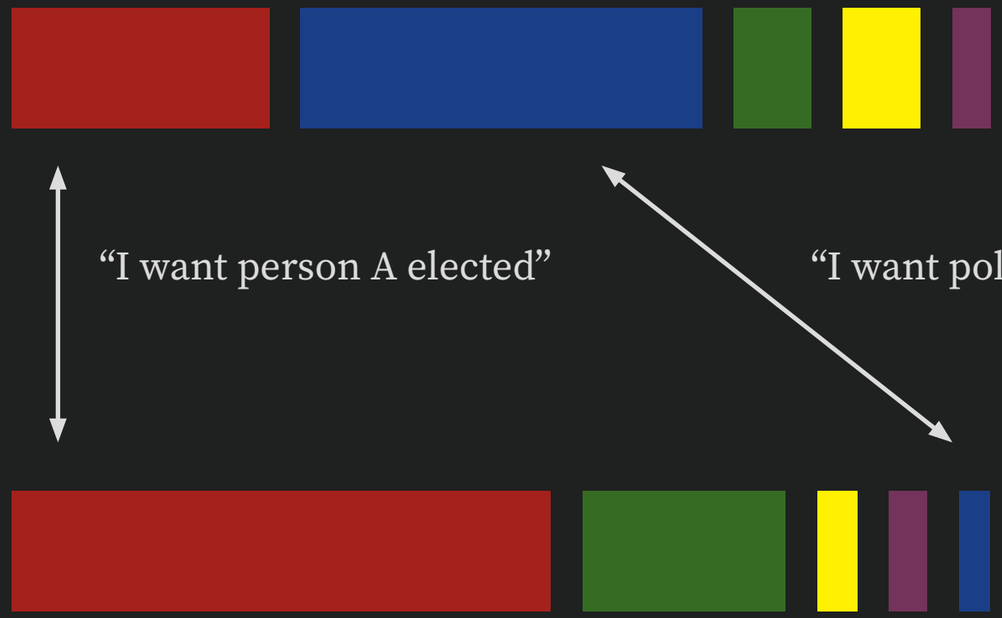
https://en.wikipedia.org/wiki/Cosine_similarity

Accounting for feature similarity

Cosine similarity is good when each feature is discrete

But we want to account for the fact that some features may be more salient than others

Accounting for feature similarity



Accounting for feature similarity

For a particular discussion, certain perspective features will be more salient (something like *relevance*)

Adjust calculation of *perspective* similarity with similarity of each *feature pair*

Use soft cosine similarity from NLP

$$\text{soft\_cosine}_1(a, b) = \frac{\sum_{i,j}^{N} s_{ij} a_i b_j}{\sqrt{\sum_{i,j}^{N} s_{ij} a_i a_j} \sqrt{\sum_{i,j}^{N} s_{ij} b_i b_j}},$$

https://en.wikipedia.org/wiki/Cosine_similarity

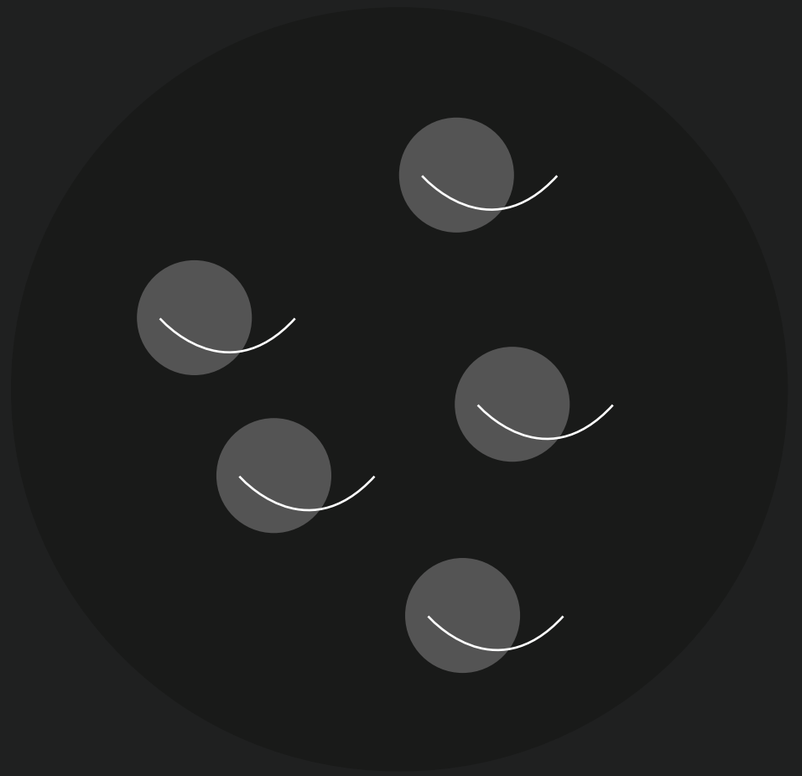
Primary metric: group homogeneity

Homogeneity is determined by:

Number of agents

Perspective makeup of each agent

Feature relatedness



Homogeneity influences outcome

+ $G_{homogeneity}$ = outcome likely to be the most prevalent perspective

- $G_{homogeneity}$ = outcome likely more random (dependent on external factors)

Each agent can have a “tendency to compromise” randomly sampled from a normal distribution

Lower homogeneity → compromise variable has greater weight in outcome calculation

Summing up

Representing each agent's perspective as a normalized vector of features

Can use soft cosine similarity to calculate group homogeneity

Homogeneity used to calculate group outcome

Overall goal: explore how different partition methods result in different outcomes using this framework

Q&A