Applied NLP: Static word embeddings in Computational Social Science

CS 2731/ISSP 2230 guest lecture

Neha Kennard

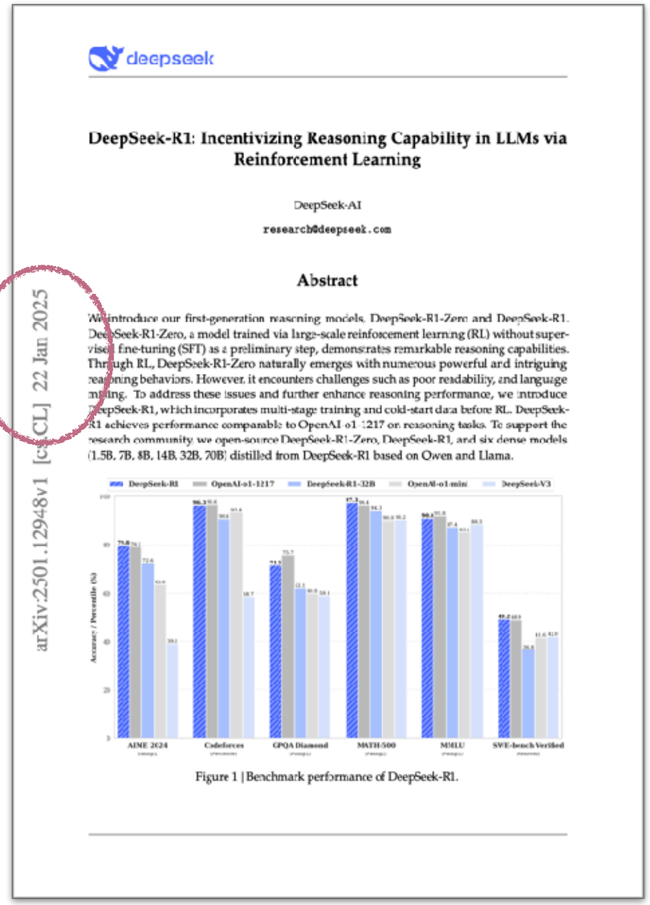
January 29 2025

A little about me

- PhD student at IESL
- UMass Amherst
- NLP for Computational Social Science
- End users: sociologists

Schedule

Date	Topics	Slides	Readings
Wednesday, Jan 8	Introduction	Course Introduction	
Monday, Jan 13	Project Ideas	Project Ideas and requirements	
Wednesday, Jan 15	Text Processing	Basic text processing	Jurafsky and Martin Chapter 2 (2.1-2.3, 2.5-2.7)
Monday, Jan 20	No Class		
Wednesday, Jan 22	Representation Le...	Sparse word representation	Jurafsky and Martin Chapter 6 (6.3-6.7)
Monday, Jan 27	Representation Le...	Dense word representation	Jurafsky and Martin Chapter 6 (6-6.2, 6.8-6.13) Language (Technology) is Power: A Critical Survey of "Bias" in NLP
Wednesday, Jan 29	Representation Le...	Guest Lecture: Neha Nayak Kennard Topic: Static word embeddings in Computational Social Science	
Monday, Feb 3	Text Classification	Logistic regression CRC Tutorial	Jurafsky and Martin Chapter 5 (5-5.3)
Wednesday, Feb 5	Text Classification	Logistic regression 2	Jurafsky and Martin Chapter 5 (5.4-5.6, 5.11)
Monday, Feb 10	Text Classification Neural Networks	Feedforward Neural Network	Jurafsky and Martin Chapter 7 (7-7.1, 7.3-7.4, 7.6, 7.8)
Wednesday, Feb 12	Text Classification	Classifier evaluation	Jurafsky and Martin Chapter 4 (4.7-4.10)
Monday, Feb 17	Project Proposal Pr...		
Wednesday, Feb 19	Language Modeling		
Monday, Feb 24	Language Modeling Neural Networks		
Wednesday, Feb 26	Language Modeling Neural Networks		
Monday, Mar 4	No Class		
Wednesday, Mar 6	No Class		
Monday, Mar 10	Language Modeling		
Wednesday, Mar 12	Language Modeling		
Monday, Mar 17	Language Modeling		
Wednesday, Mar 19	Language Modeling		
Monday, Mar 24	Sequence Labeling		
Wednesday, Mar 26	Parsing		
Monday, Mar 31	Reasoning in NLP		DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via



deepseek

DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning

DeepSeek-AI
research@deepseek.com

Abstract

We introduce our first-generation reasoning models, DeepSeek-R1-Zero and DeepSeek-R1. DeepSeek-R1-Zero, a model trained via large-scale reinforcement learning (RL) without supervised fine-tuning (SFT) as a preliminary step, demonstrates remarkable reasoning capabilities. Through RL, DeepSeek-R1-Zero naturally emerges with numerous powerful and intriguing reasoning behaviors. However, it encounters challenges such as poor readability, and language mixing. To address these issues and further enhance reasoning performance, we introduce DeepSeek-R1, which incorporates multi-stage training and cold-start data before RL. DeepSeek-R1 achieves performance comparable to OpenAI-o1-1217 on reasoning tasks. To support the research community, we open-source DeepSeek-R1-Zero, DeepSeek-R1, and six dense models (1.5B, 7B, 8B, 14B, 32B, 70B) distilled from DeepSeek-R1 based on Qwen and Llama.

Figure 1 | Benchmark performance of DeepSeek-R1.

DeepSeek's Rise: How a Chinese Start-Up Went From Stock Trader to A.I. Star

Here's what DeepSeek AI does better than OpenAI's ChatGPT

How China's DeepSeek Outsmarted America

DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via

Schedule ···

Date	Topics	Slides	Readings
Wednesday, Jan 8	Introduction	Course Introduction	
Monday, Jan 13	Project Ideas	Project Ideas and requirements	
Wednesday, Jan 15	Text Processing	Basic text processing	Jurafsky and Martin Chapter 2 (2.1-2.3, 2.5-2.7)
Monday, Jan 20	No Class		
Wednesday, Jan 22	Representation Le...	Sparse word representation	Jurafsky and Martin Chapter 6 (6.3-6.7)
Monday, Jan 27	Representation Le...	Dense word representation	Jurafsky and Martin Chapter 6 (6-6.2, 6.8-6.13) Language (Technology) is Power: A Critical Survey of "Bias" in NLP
Wednesday, Jan 29	Representation Le...	Guest Lecture: Neha Nayak Kennard Topic: Static word embeddings in Computational Social Science	
Monday, Feb 3	Text Classification	Logistic regression CRC Tutorial	Jurafsky and Martin Chapter 5 (5-5.3)
Wednesday, Feb 5	Text Classification	Logistic regression 2	Jurafsky and Martin Chapter 5 (5.4-5.6, 5.11)
Monday, Feb 10	Text Classification Neural Networks	Feedforward Neural Network	Jurafsky and Martin Chapter 7 (7-7.1, 7.3-7.4, 7.6, 7.8)
Wednesday, Feb 12	Text Classification	Classifier evaluation	Jurafsky and Martin Chapter 4 (4.7-4.10)
Monday, Feb 17	Project Proposal Pr...		
Wednesday, Feb 19	Language Modeling		
Monday, Feb 24	Language Modeling Neural Networks		
Wednesday, Feb 26	Language Modeling Neural Networks		
Monday, Mar 4	No Class		
Wednesday, Mar 6	No Class		
Monday, Mar 10	Language Modeling		
Wednesday, Mar 12	Language Modeling		
Monday, Mar 17	Language Modeling		
Wednesday, Mar 19	Language Modeling		
Monday, Mar 24	Sequence Labeling		
Wednesday, Mar 26	Parsing		
Monday, Mar 31	Reasoning in NLP		DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via

Dense word representation

Showing 1,339 results for:
chatgpt

The New York Times

GIVE THE TIMES

Showing 1 results for:
word2vec

Sort by Relevance Search in English

Date Range

Aug. 9, 2016

TECHNOLOGY

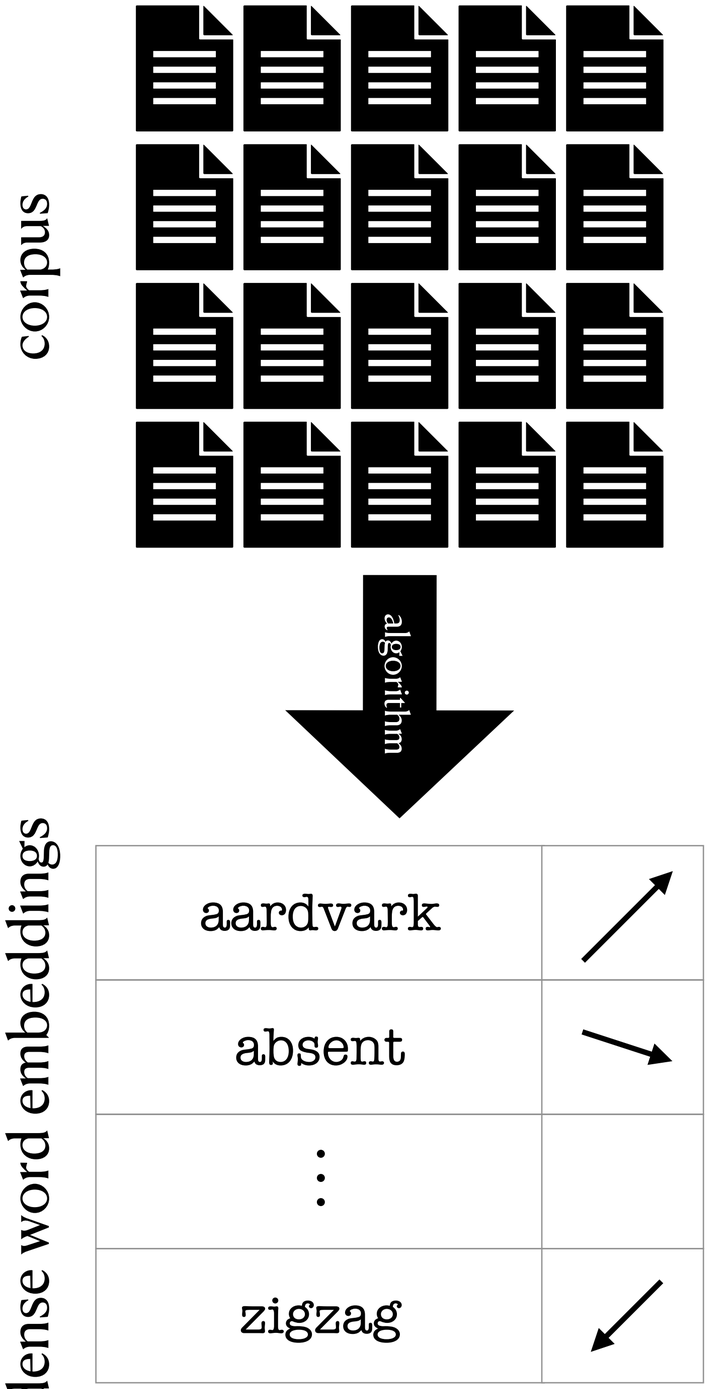
Summer Intern Report: Prototyping an Improved Search Query With Machine Learning

One of our Search interns shares details about his summer project — improving word relevancy to increase the relevancy of search results — and his experience working with the team.

By Tushar Baraiya

Questions for today

- Who is still using static word representations, and for what?
- What can I, a computer scientist, contribute?
- Why not just use LLMs instead?



What are static word embeddings?

ICYMI

- Given a corpus of text(s)
- An algorithm such as word2vec, GloVe, fasttext... is applied
- Producing a set of vectors (points in a vector space)
- One for each unique word (‘type’) that occurs in the corpus
- Similarity functions apply to pairs of vectors (e.g. cosine, Euclidean distance)

$sim(\mathbf{w_1}, \mathbf{w_2})$ HIGH ⟷ similar contexts ⟷ similar meanings

⇕

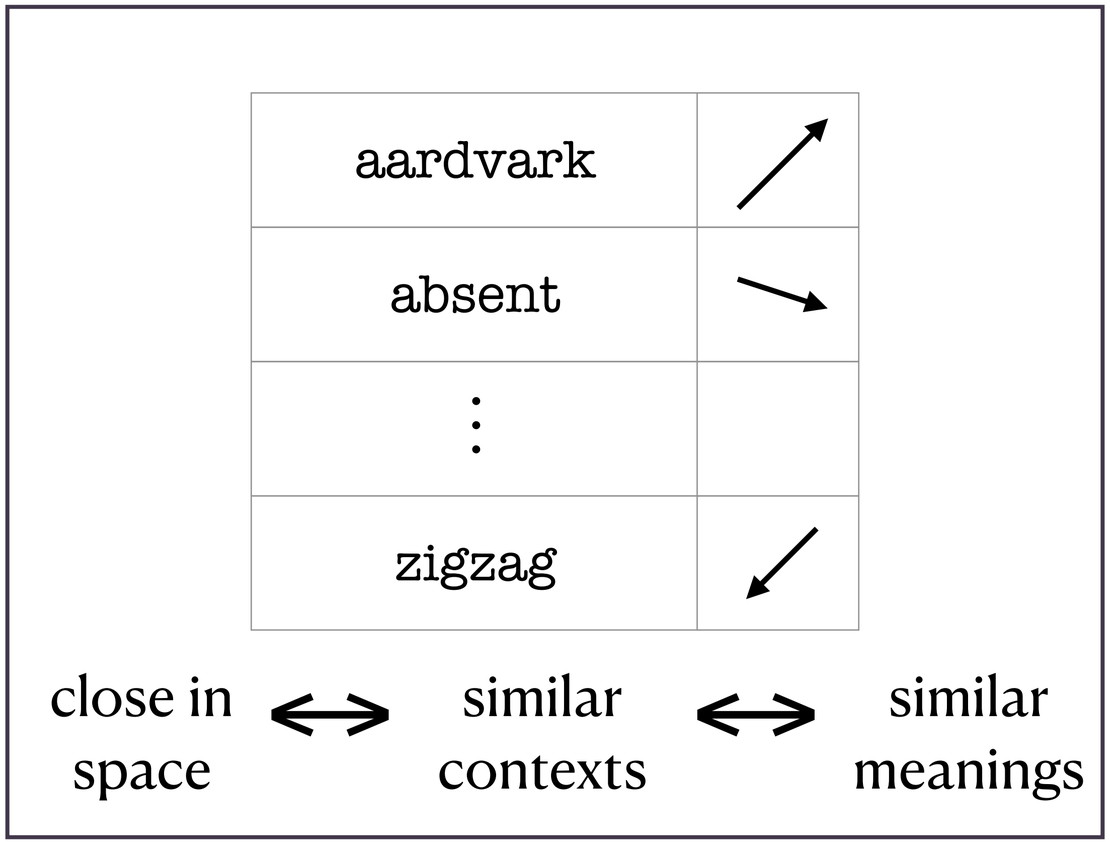
close in space

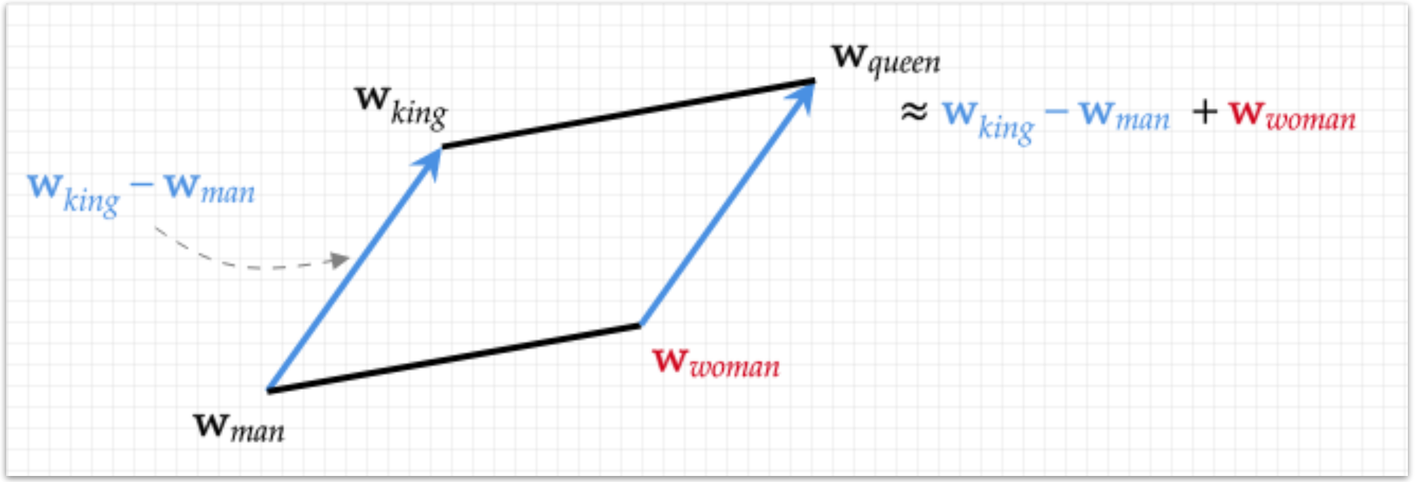
"You shall know a word by the company it keeps" - [Firth 1957]

$sim(\mathbf{w_1}, \mathbf{w_2})$ LOW ⟷ dissimilar contexts ⟷ dissimilar meanings

⇕

far apart


aardvark
absent
zigzag
close in space
similar contexts
similar meanings


$\mathbf{w}_{queen}$
$\approx \mathbf{w}_{king} - \mathbf{w}_{man} + \mathbf{w}_{woman}$
$\mathbf{w}_{king}$
$\mathbf{w}_{king} - \mathbf{w}_{man}$
$\mathbf{w}_{woman}$
$\mathbf{w}_{man}$

Questions for today

- Who is still using static word representations, and for what?
- What can I, a computer scientist, contribute?
- Why not just use LLMs instead?

Research questions

In NLP and in other fields

NLP:

(From CS3730)

- Can models learn language without embodiment?
- Should knowledge be neuralized or indexed?
- How will the understanding of language benefit multi-modal applications and embodied agents?

Research questions

In NLP and in other fields

NLP:
(From CS3730)

- Can **models** learn language without embodiment?
- Should knowledge be neuralized or indexed **[in NLP models]**?
- How will the understanding of language benefit multi-modal applications **[of NLP models]** and embodied **agents**?

Questions *about* NLP models

NLP models are the *subject*

Research questions

In NLP and in other fields

Political science

- How, and by whom, is emotional language employed in US Congress debates?

Question *about* people and their interactions

NLP models are a *tool*; people are the *subject*

How, and by whom, is emotional language employed in US Congress debates?

Gloria Gennaro, Elliott Ash, Emotion and Reason in Political Language, *The Economic Journal*, Volume 132, Issue 643, April 2022, Pages 1037–1059, https://doi.org/10.1093/ej/ueab104

How, and by whom, is emotional language employed in US Congress debates?

- In his treatise on Rhetoric, Aristotle suggested that persuasion can be achieved through either

logical argumentation ⚙️

or

emotional arousal 🎭

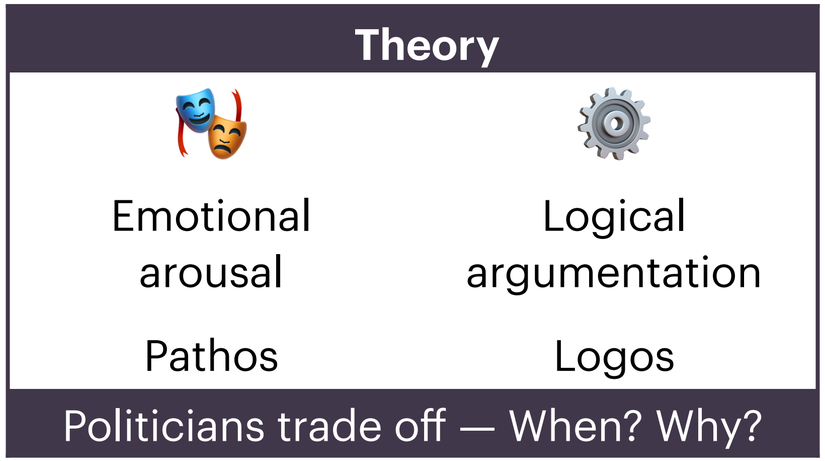
in the audience;

- success depends on selecting the most appropriate strategy for the given context.

How, and by whom, is emotional language employed in US Congress debates?

- In his treatise on Rhetoric, Aristotle suggested that persuasion can be achieved through **either logical argumentation ⚙️ or emotional arousal 🎭** in the audience; success depends on selecting the most appropriate strategy for the given context.
- The extent to which politicians engage with this trade-off … **is largely unknown**.
- **Providing empirical evidence on these questions has been difficult** due to the lack of a reproducible, validated and scalable measure of emotionality in political language.

How, and by whom, is emotional language employed in US Congress debates?



The toolkit

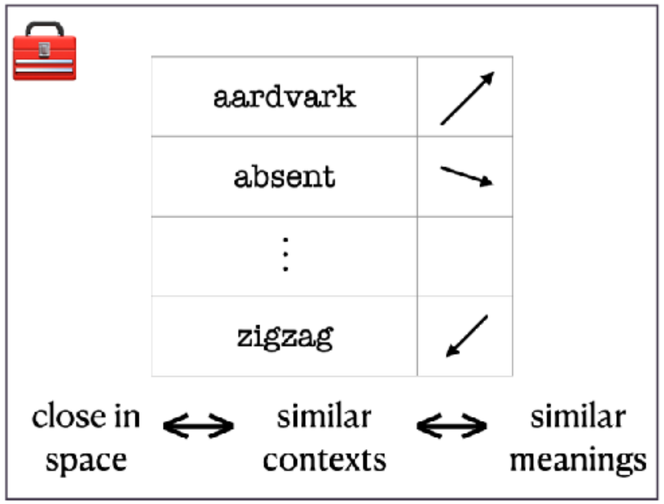
How, and by whom, is emotional language employed in US Congress debates?

- Digitized transcripts of speeches in the U.S. House and Senate between 1858 and 2014
- For each speech:
  - Full text
  - Date of speech
  - Speaker's political party

Theory

Emotional arousal	Logical argumentation
Pathos	Logos

Politicians trade off — When? Why?

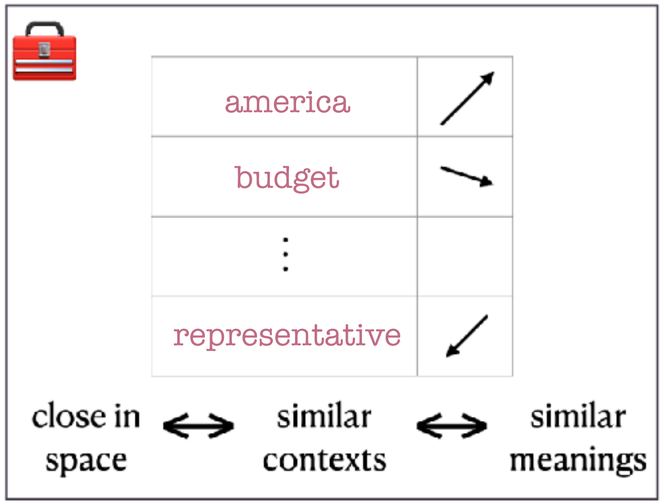


Preparation

How, and by whom, is emotional language employed in US Congress debates?

- Concatenate all speeches from 1858 - 2014 into one corpus
- Clean data (part-of-speech tagging, removing stopwords, etc.)
- Train a word2vec model

Theory	
🎭	⚙️
Emotional arousal	Logical argumentation
Pathos	Logos
Politicians trade off — When? Why?	



Method

How, and by whom, is emotional language employed in US Congress debates?

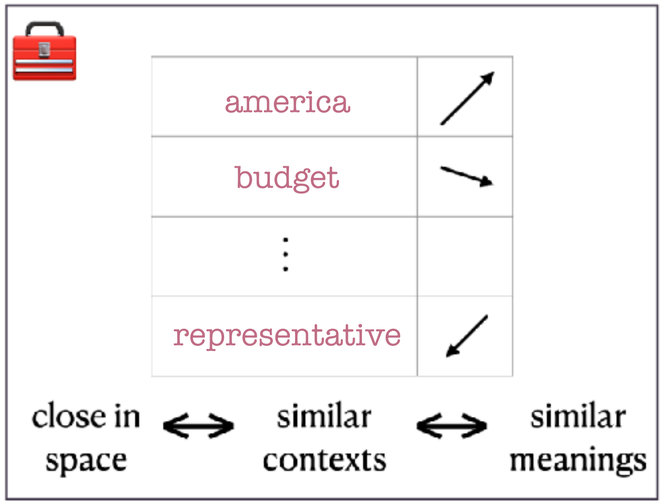
- Representations of speeches?
- Speech ↔ set of words
- Speech representation ↔ mean of vectors in set
- Why?

- set of words in speech → mean of their vectors
  → speech vector

Theory

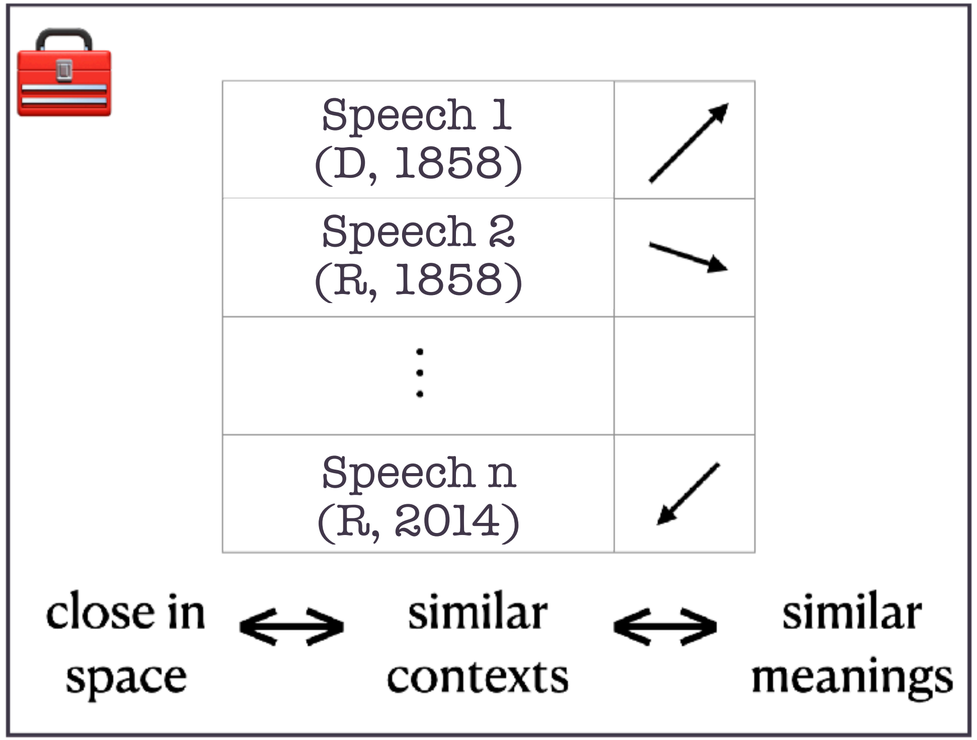
🎭 Emotional arousal	⚙️ Logical argumentation
Pathos	Logos

Politicians trade off — When? Why?



Method

How, and by whom, is emotional language employed in US Congress debates?

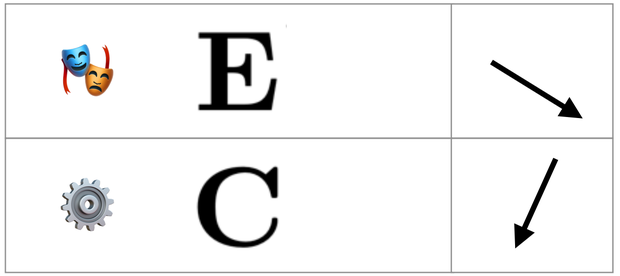


Theory	
Emotional arousal	Logical argumentation
Pathos	Logos
Politicians trade off — When? Why?	

We report there the affect dictionary words with their count in the corpus:

support (1765047), import (1421018), like (1327182), great (1195251), agre (1147658), care (1018579), help (945406), concern (834363), thank (746428), opportun (662106), defens (647623), polit (560160), interest (511530), critic (358826), credit (355314), favor (344079), open (330082), give (312834), person (297694), valu (295900), fight (273278), encourag (255137), fail (254356), relief (244541), argument (234996), attack (231244),

set of concept words → construct mean of their vectors → concept vector

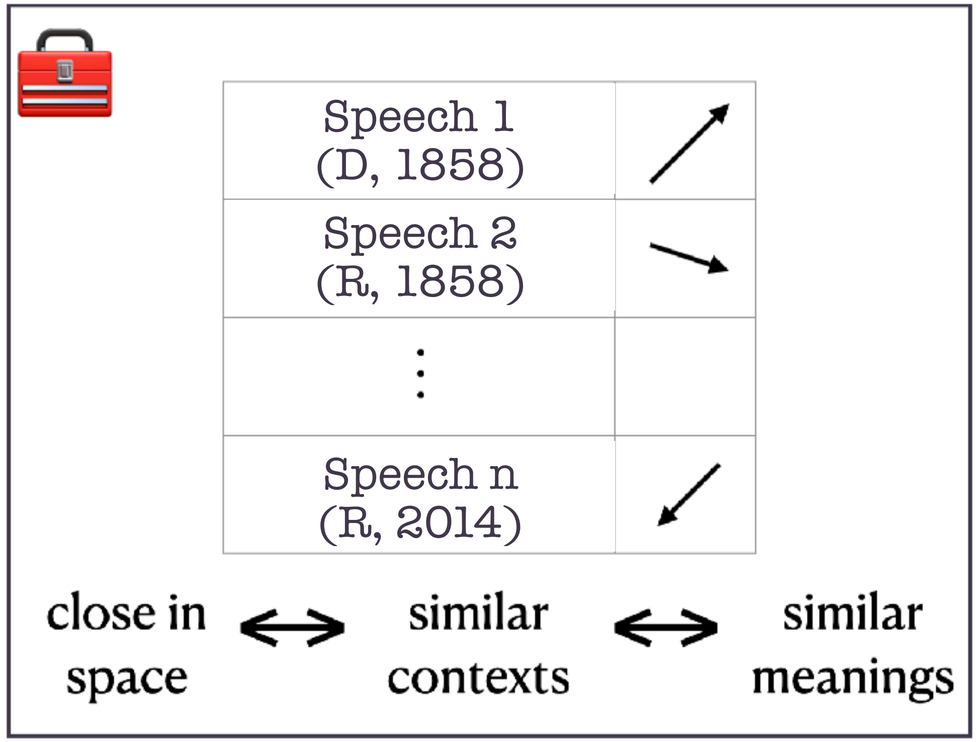


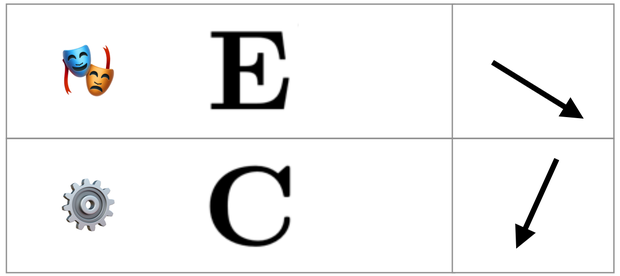
We report there the cognition dictionary words with their count in the corpus:

think (2222390), want (1933090), need (1858735), question (1765467), know (1761052), believ (1294547), fact (1278946), resolut (1204296), reason (870024), understand (860049), effect (829068), consid (802972), chang (800344), purpos (794236), make (755361), allow (741097), product (738070), recogn (722642), result (685842), control (675044), distinguish (672218), respons (669281), statement (649465), inform (628884), differ (616581), refer

Method

How, and by whom, is emotional language employed in US Congress debates?





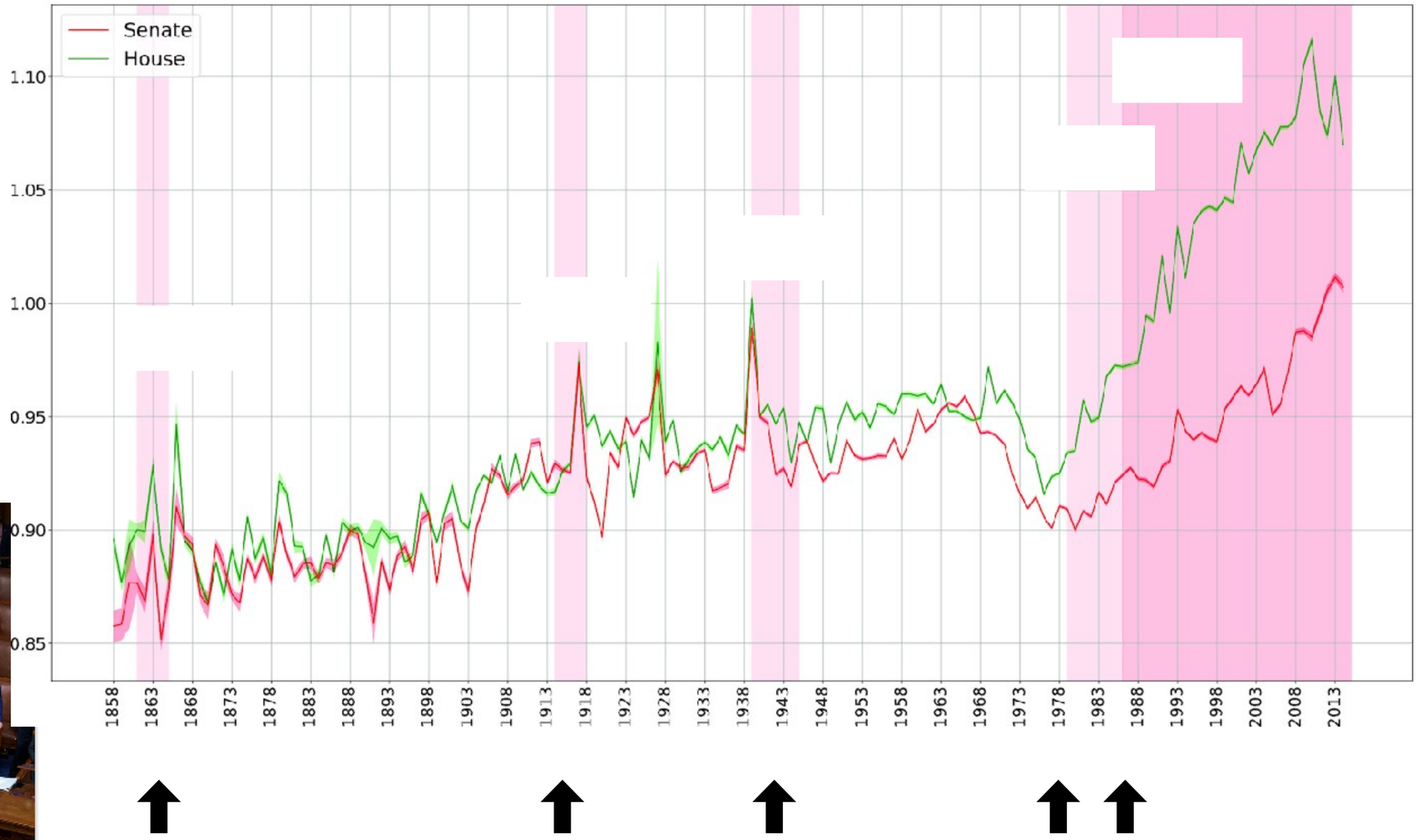
- Speech *i* is represented by vector $\mathbf{d}_i$ — mean of vectors of all words in the speech
- *Emotionality* Y_i of speech *i*

$$Y_i = \frac{sim(\mathbf{d}_i, \mathbf{E}) + b}{sim(\mathbf{d}_i, \mathbf{C}) + b}$$

Findings

How, and by whom, is emotional language employed in US Congress debates?

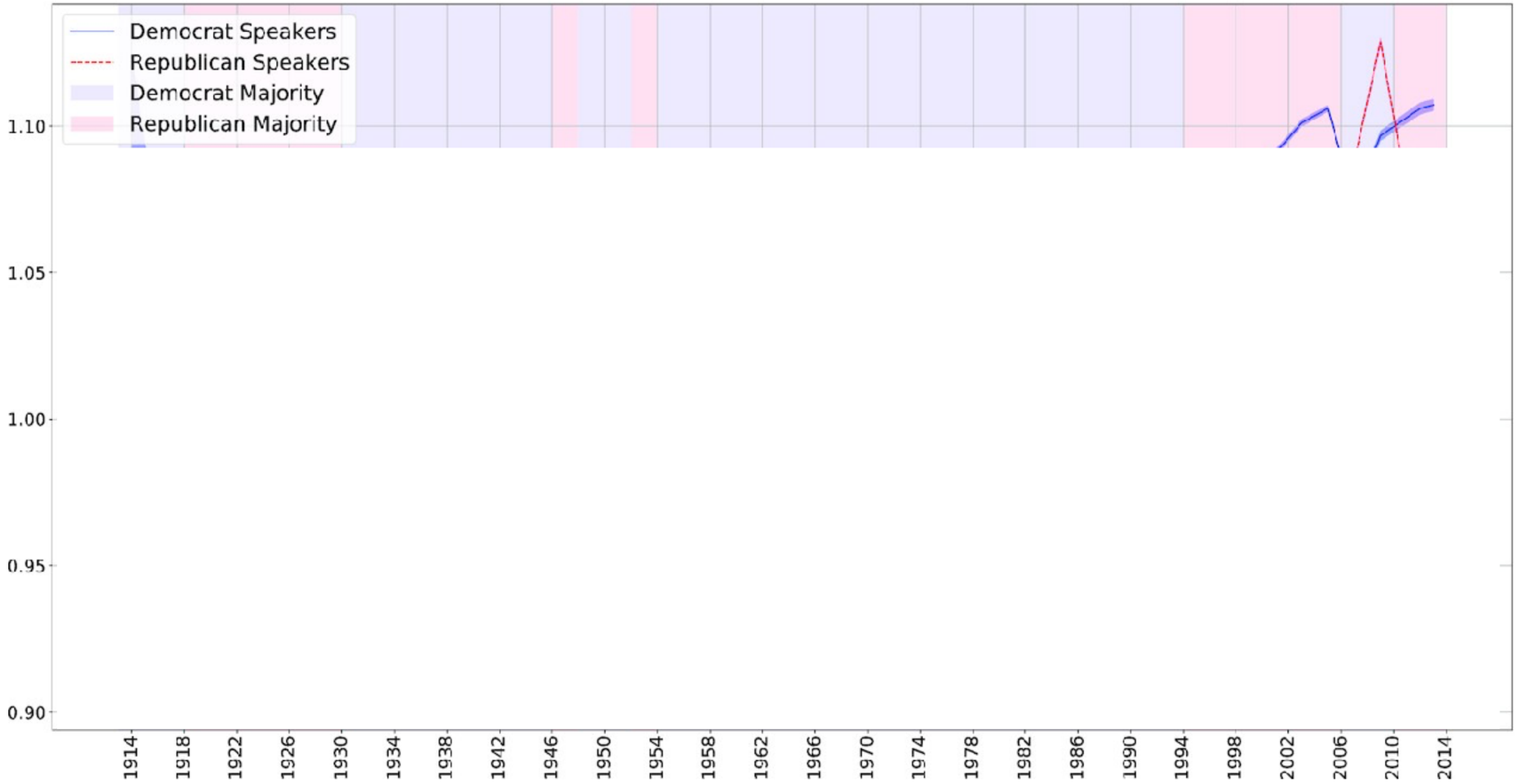
- Emotionality over time



Findings

How, and by whom, is emotional language employed in US Congress debates?

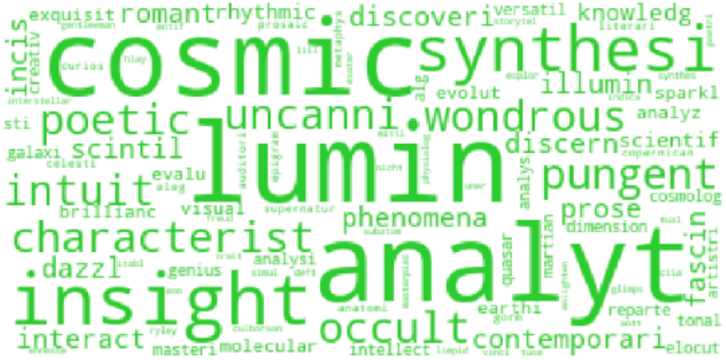
- Emotionality by party and party majority



Double checking

How, and by whom, is emotional language employed in US Congress debates?

- Could this measure accidentally be measuring positive v/s negative sentiment?
  - No! They run the same experiment with positive and negative words, and find that they are *not correlated*



(a) Cognitive Positive Language

(b) Emotional Positive Language

(c) Cognitive Negative Language

(d) Emotional Negative Language

Double checking

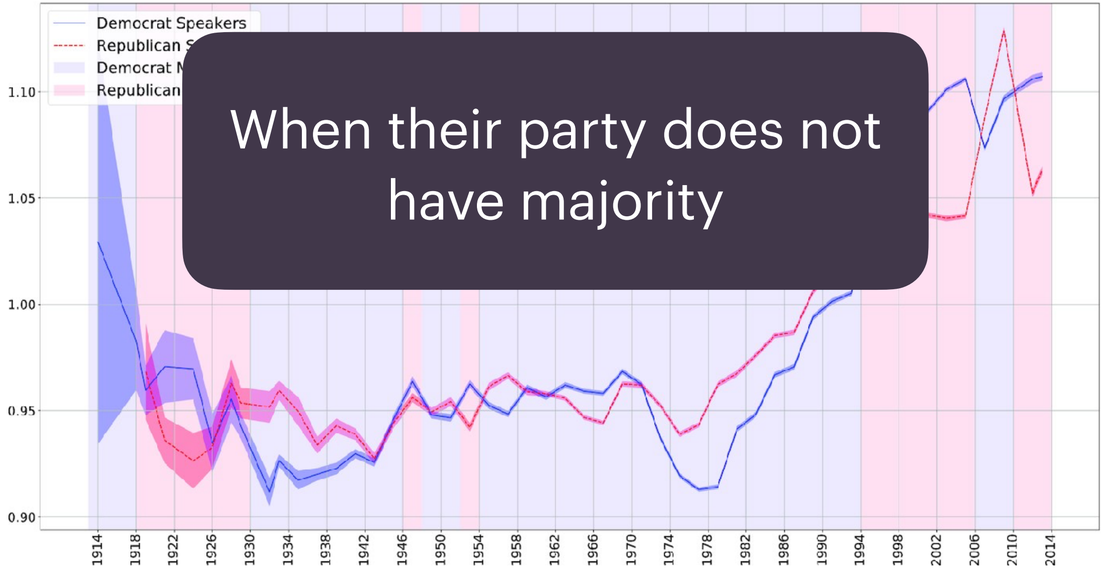
How, and by whom, is emotional language employed in US Congress debates?

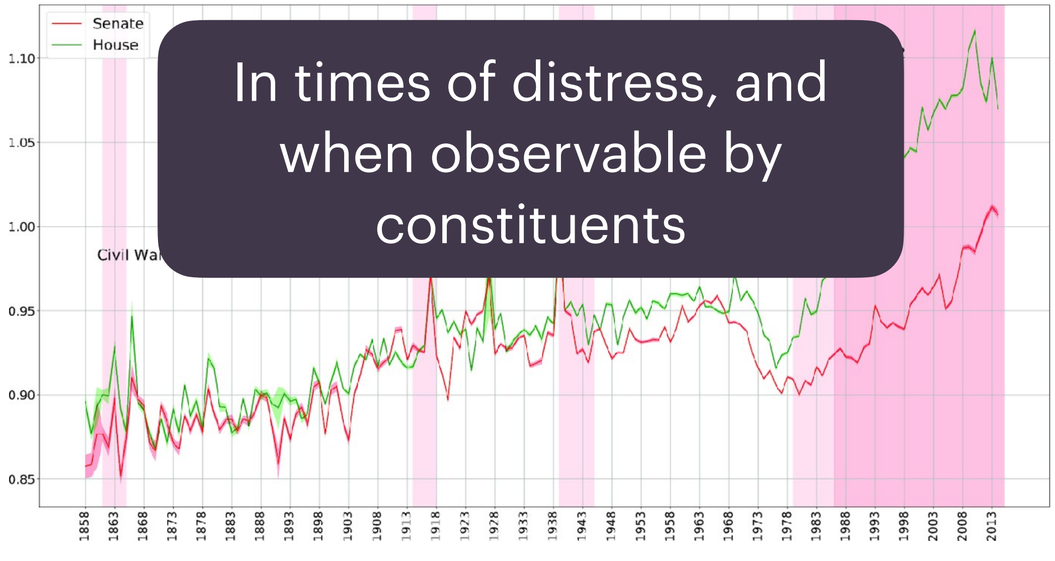
- Is this general language change, rather than something specific occurring in politics?
  - No! They run the same experiment for Google Books and find emotionality *decreasing*
- Is this the same as polarization? (Different parties gravitating to different topics)
  - No! Prior work has found polarization, but *starting in the 1990s*

Theory	
Emotional arousal	Logical argumentation
Pathos	Logos
Politicians trade off — When? Why?	

Conclusion

- The extent to which politicians engage with this trade-off … **is largely unknown**.
- But, in the US Congress, we can say quantitatively using **static word embeddings** that politicians employ emotional language…

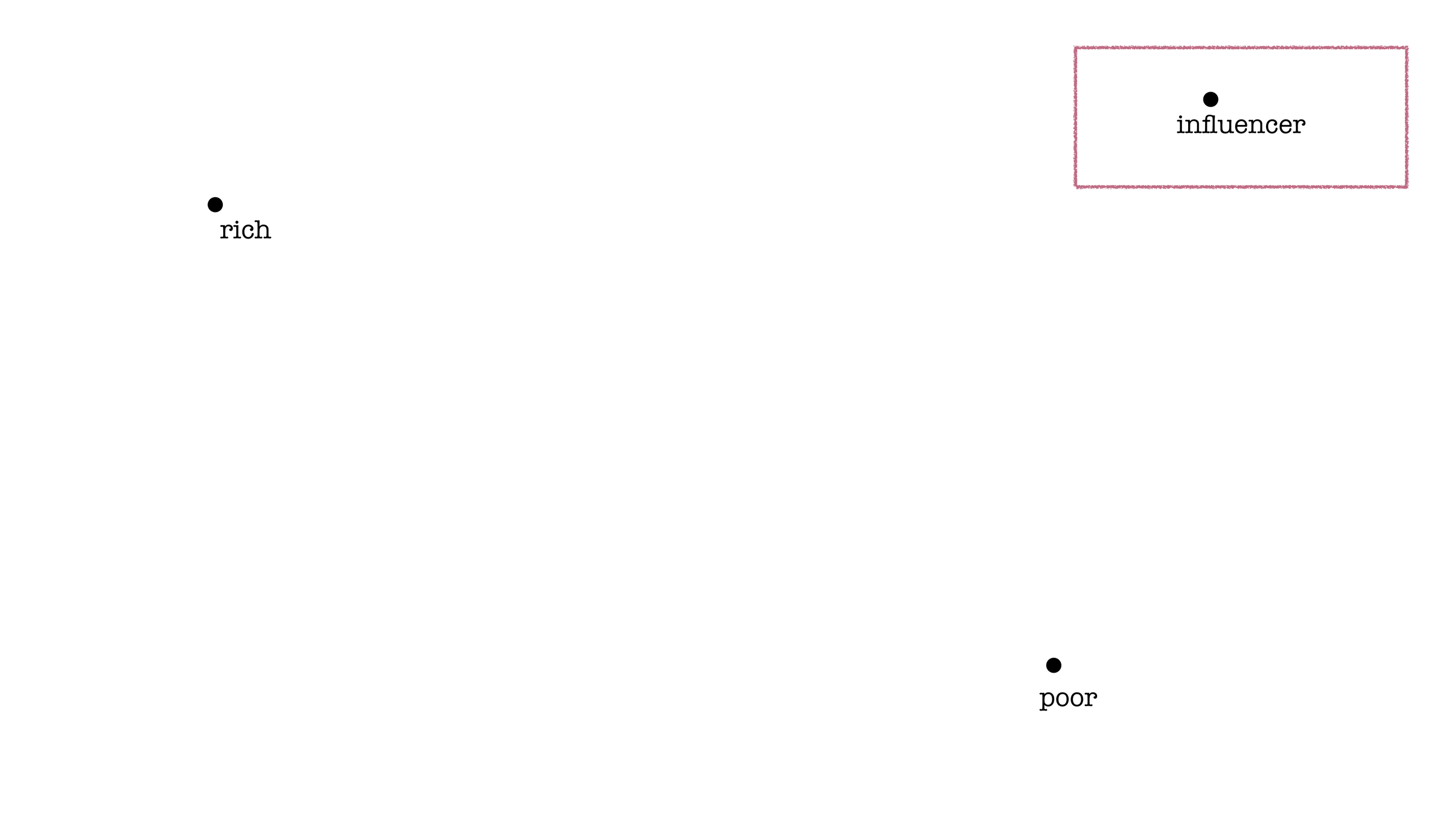




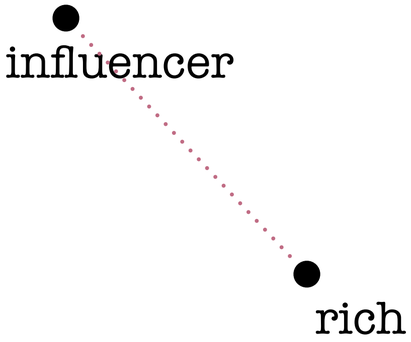
Gloria Gennaro, Elliott Ash, Emotion and Reason in Political Language, *The Economic Journal*, Volume 132, Issue 643, April 2022, Pages 1037–1059, https://doi.org/10.1093/ej/ueab104

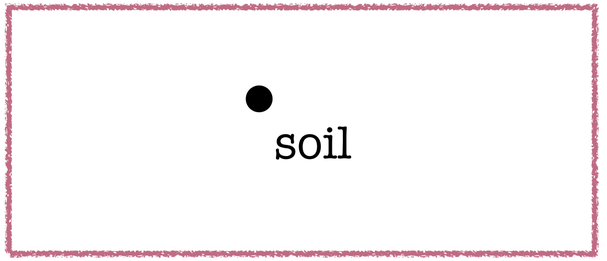
Once there's a working GitHub repo out there, are computer scientists of any use?

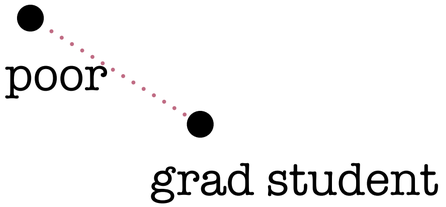
Scenario

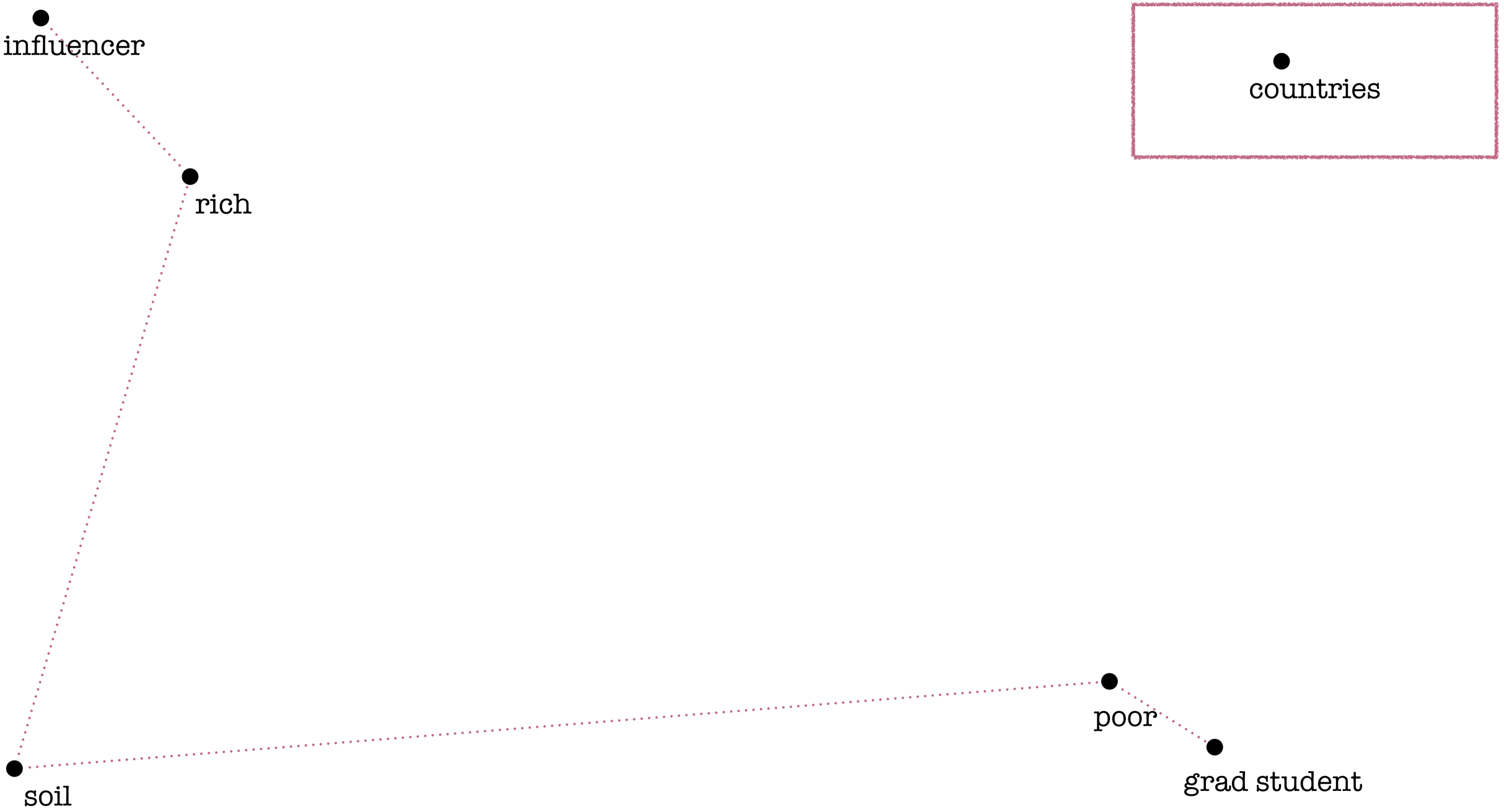

influencer
rich
poor

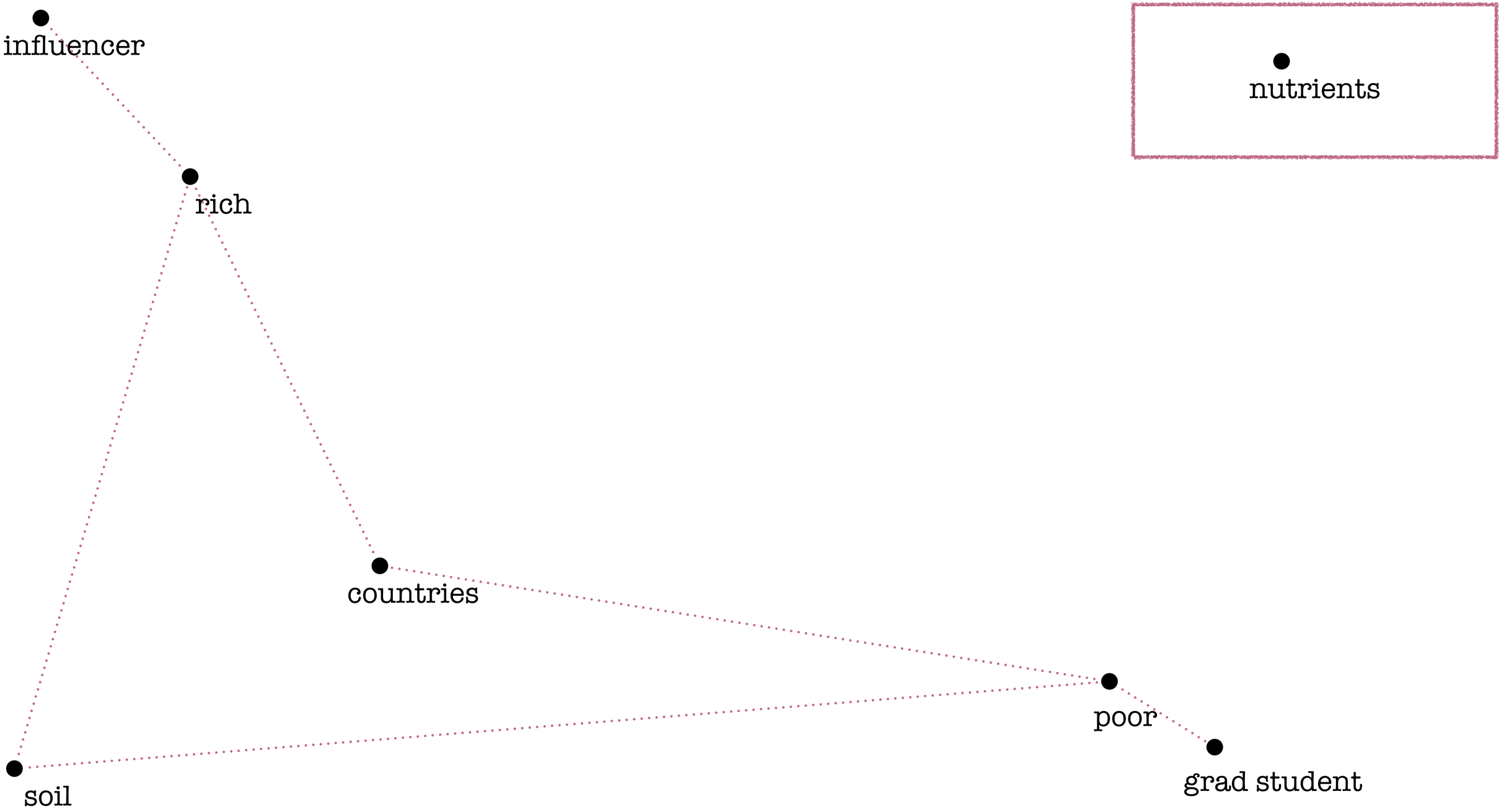

influencer
rich
grad student
poor

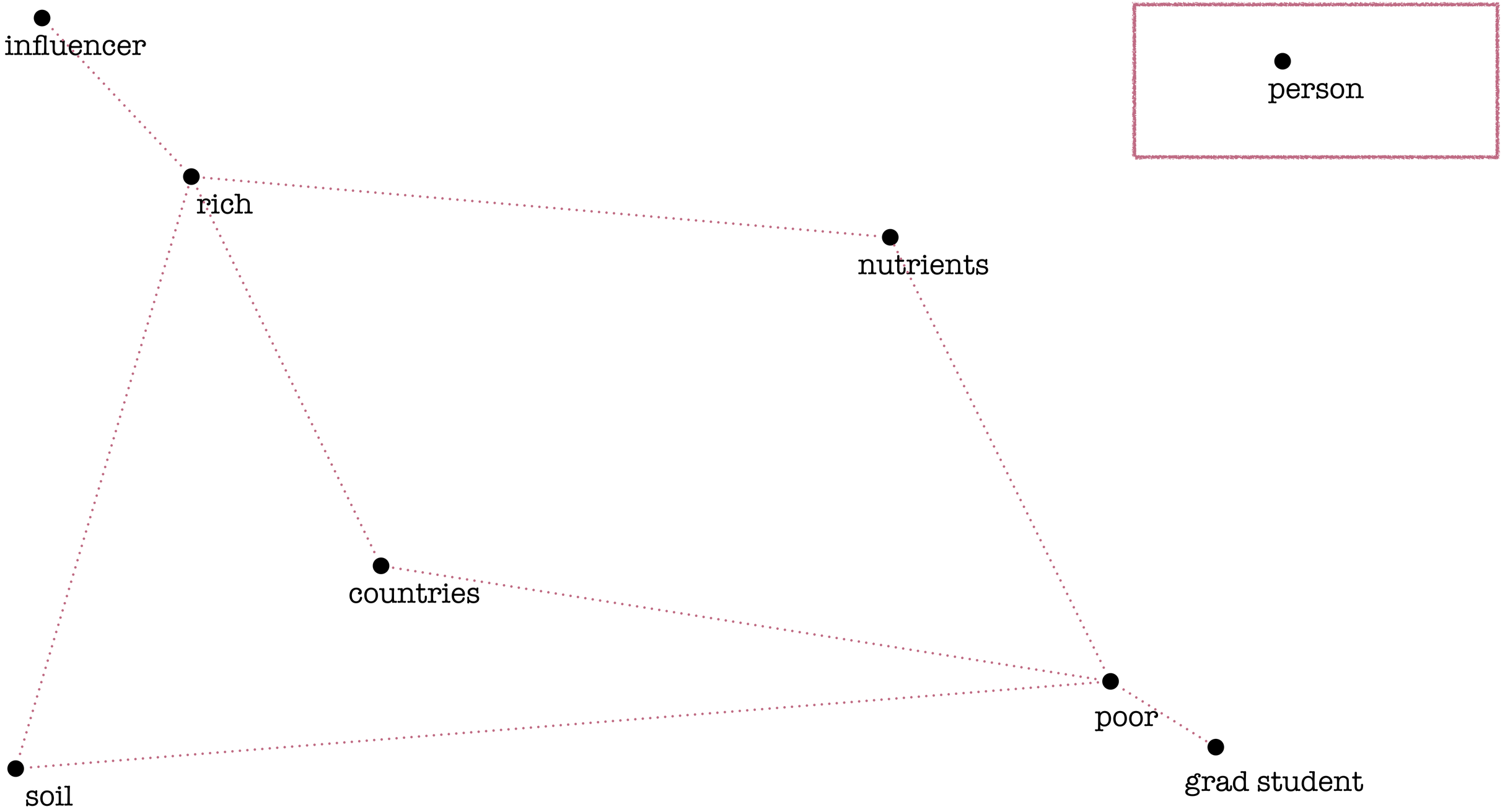

influencer
rich

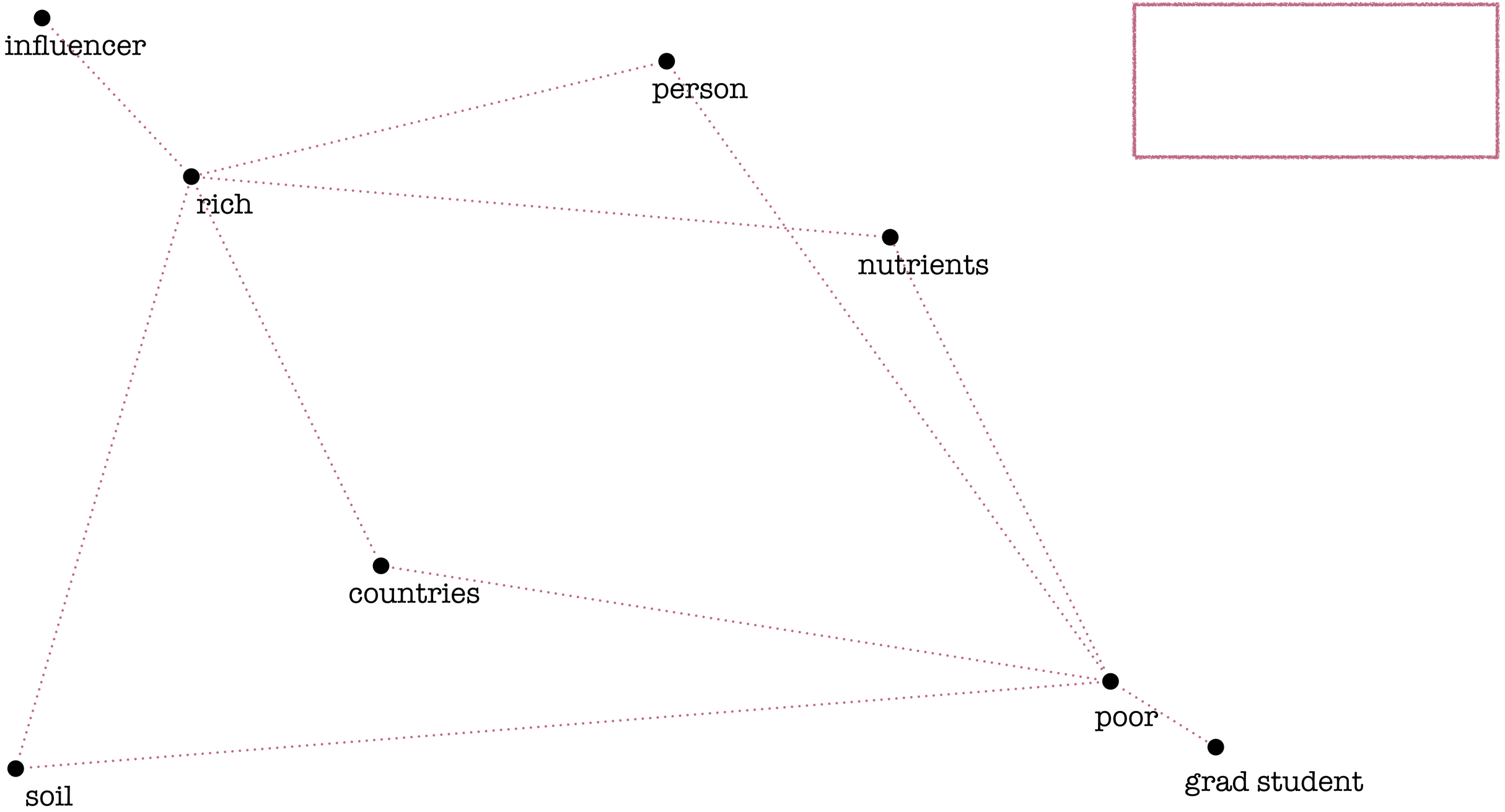

soil

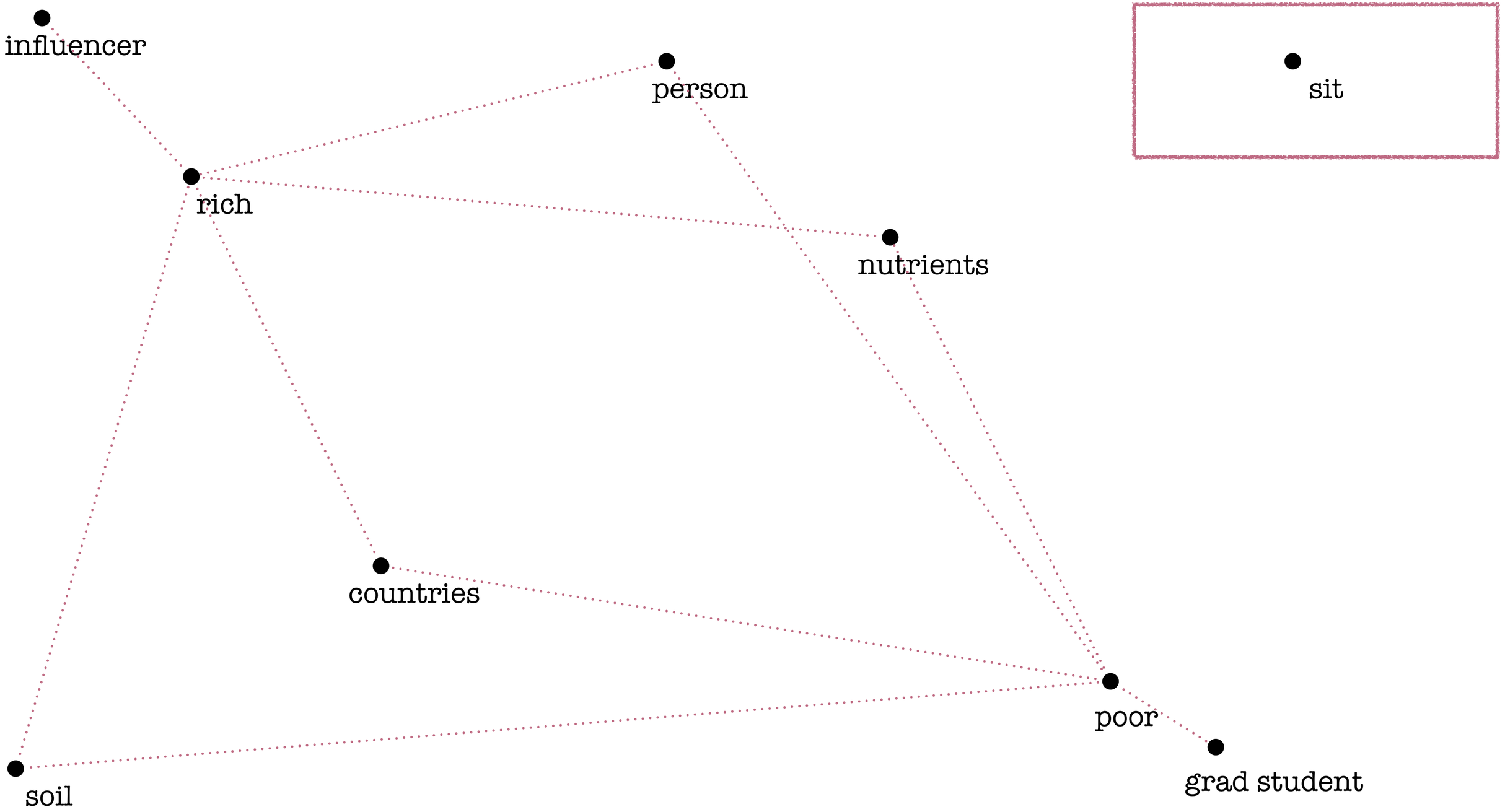

poor
grad student

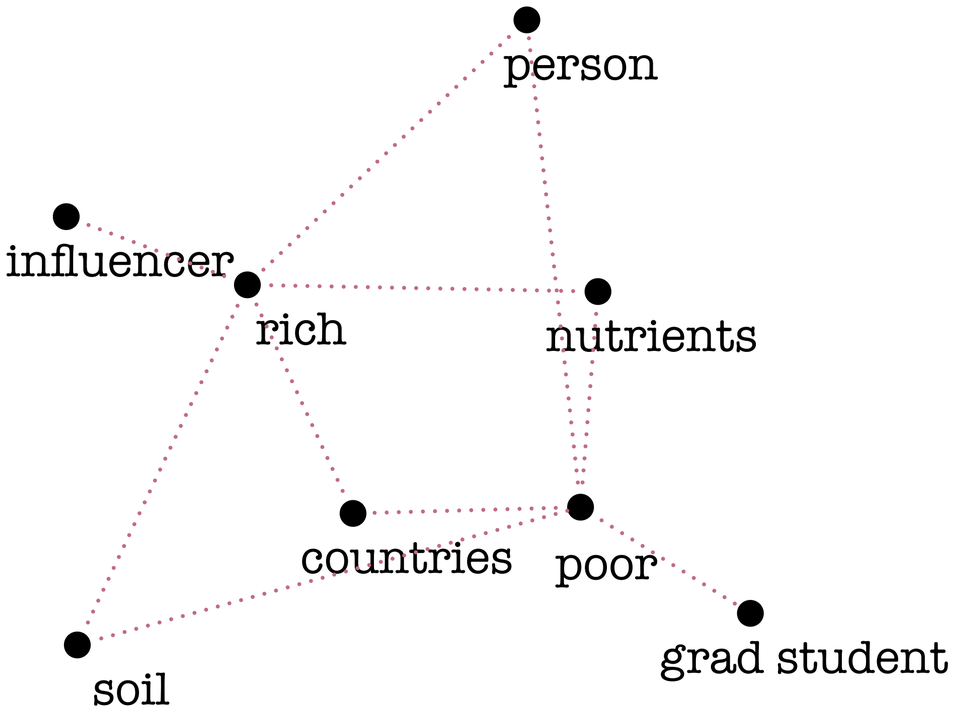

influencer
rich
countries
soil
poor
grad student

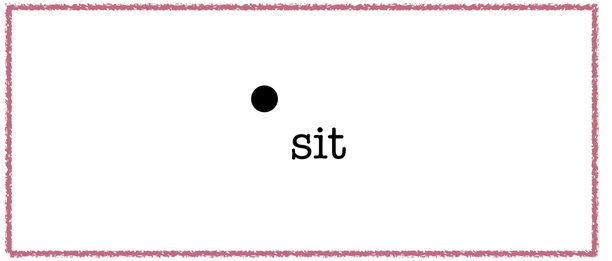

influencer
rich
countries
soil
poor
grad student
nutrients

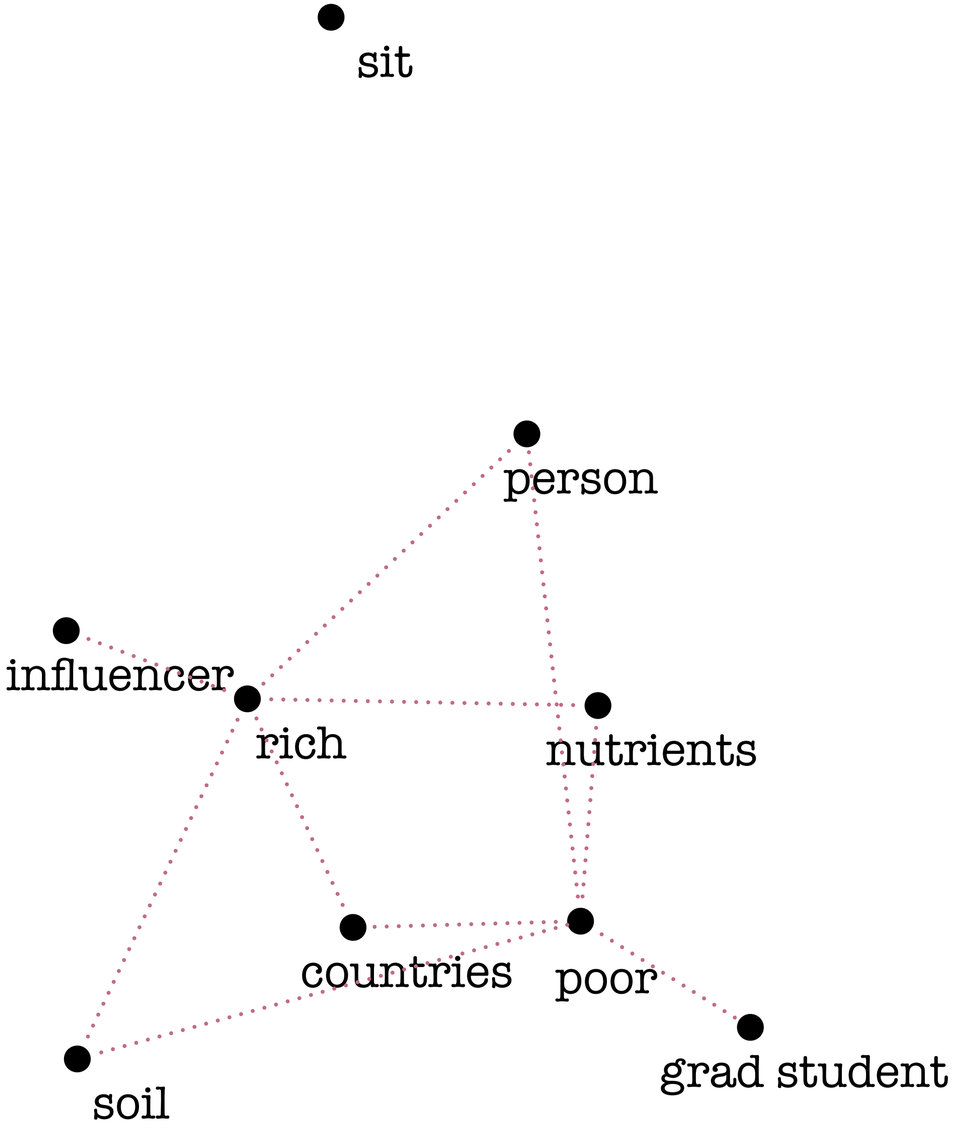

influencer
rich
nutrients
countries
poor
grad student
soil
person


influencer
person
rich
nutrients
countries
poor
soil
grad student

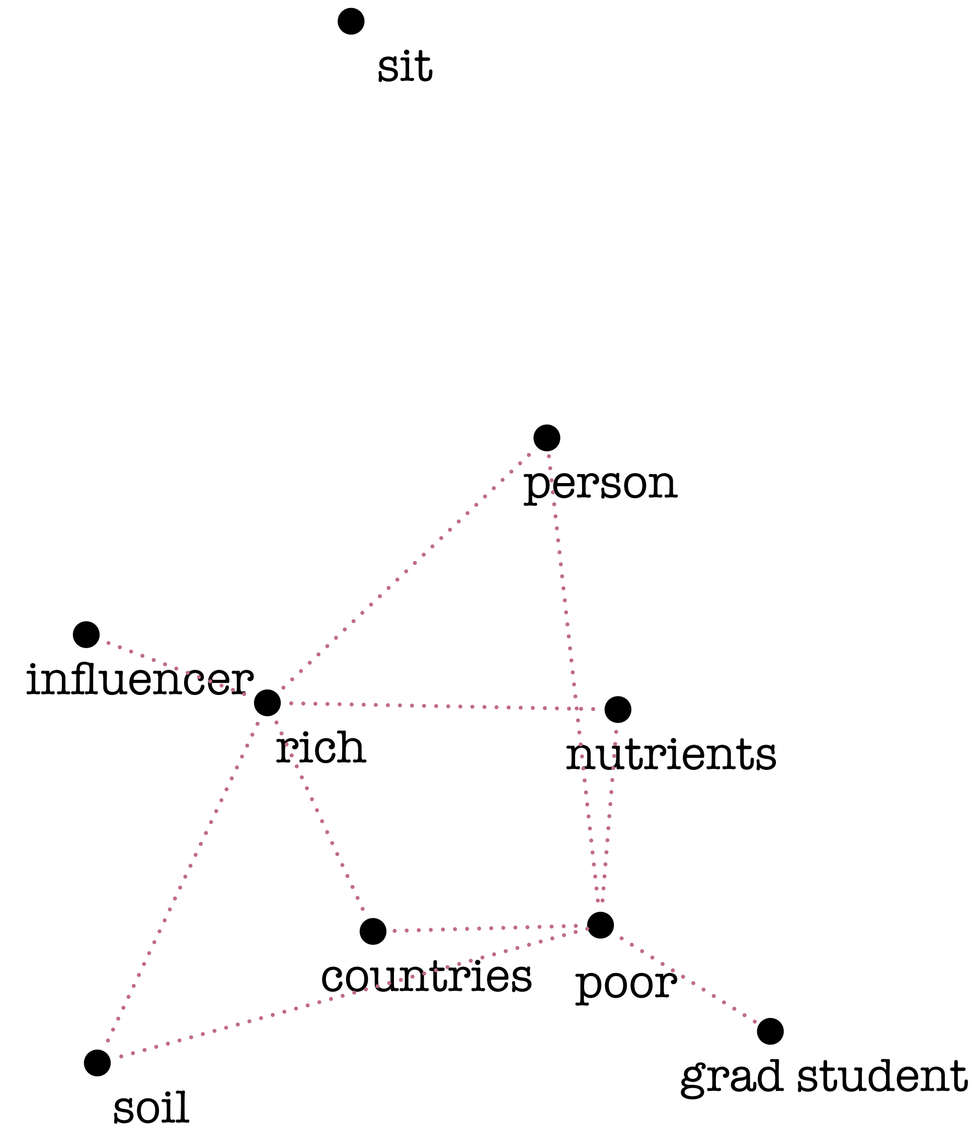

influencer
person
sit
rich
nutrients
countries
poor
soil
grad student

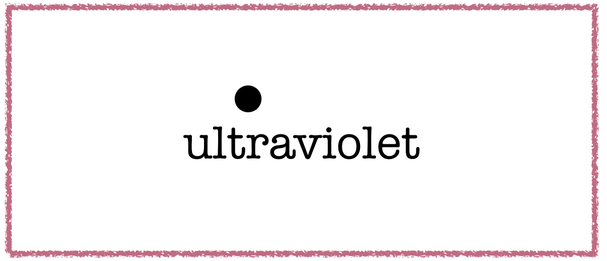

person
influencer
rich
nutrients
countries
poor
soil
grad student


sit

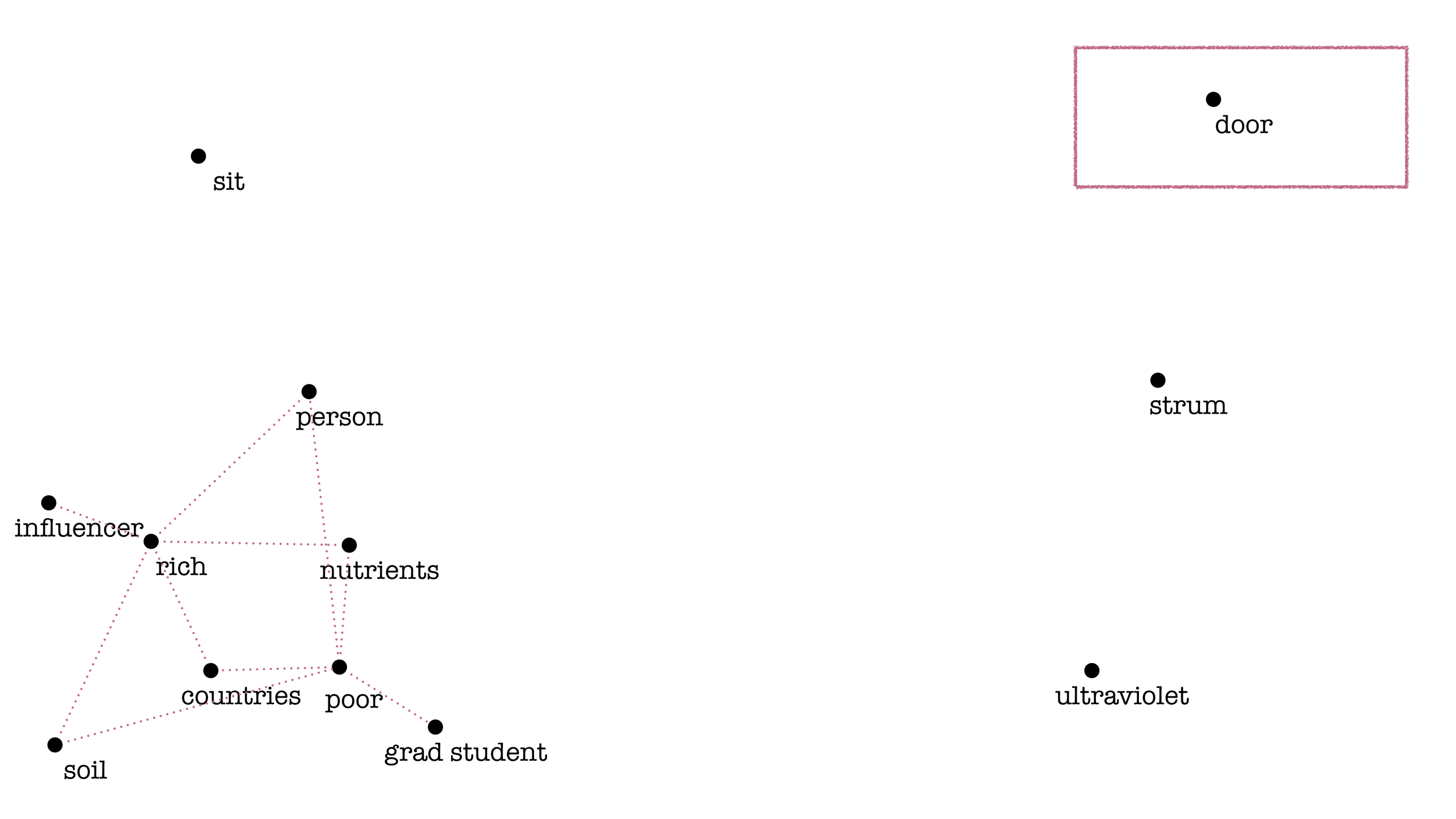

sit
person
influencer
rich
nutrients
countries
poor
soil
grad student

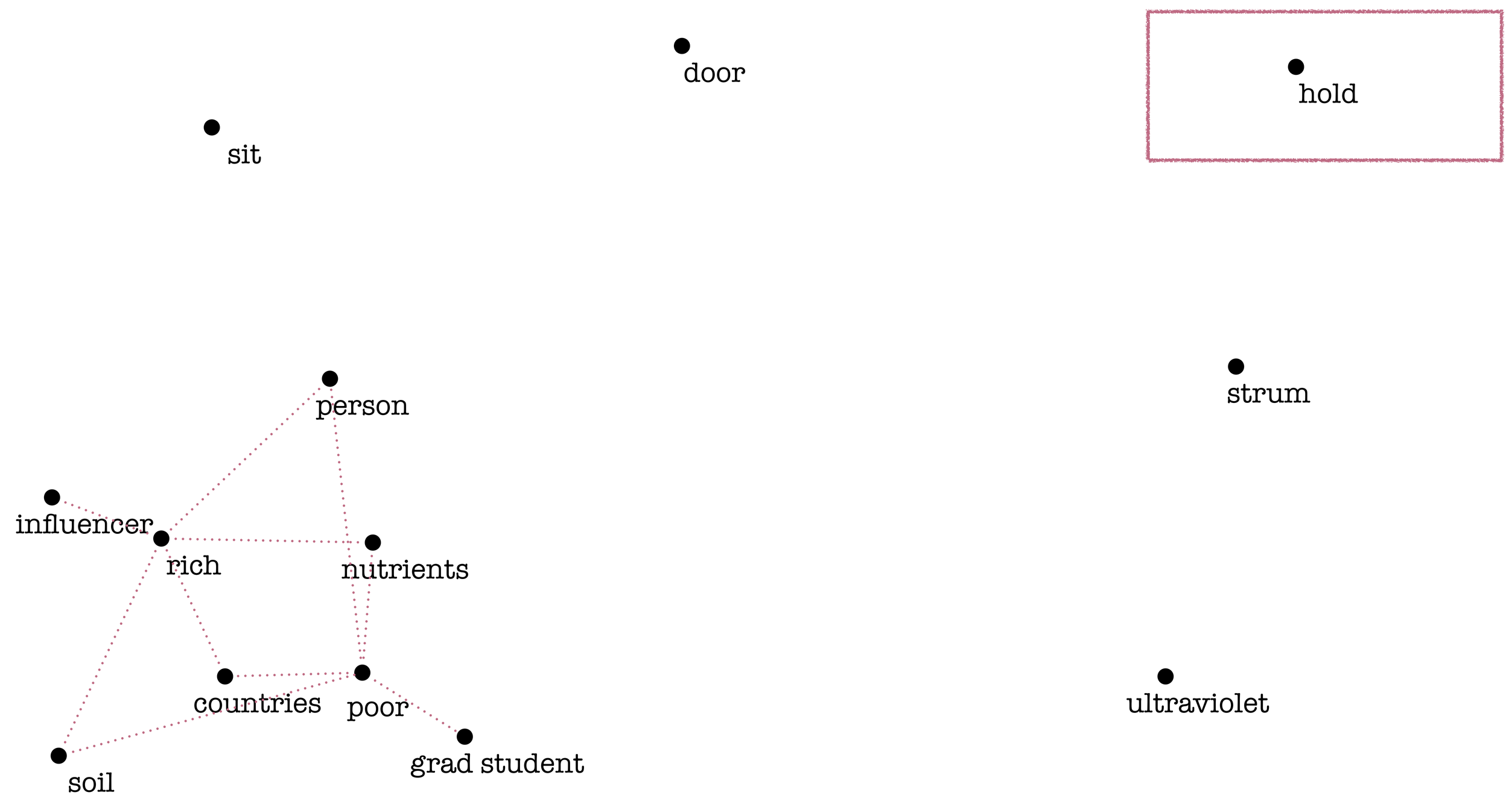
strum

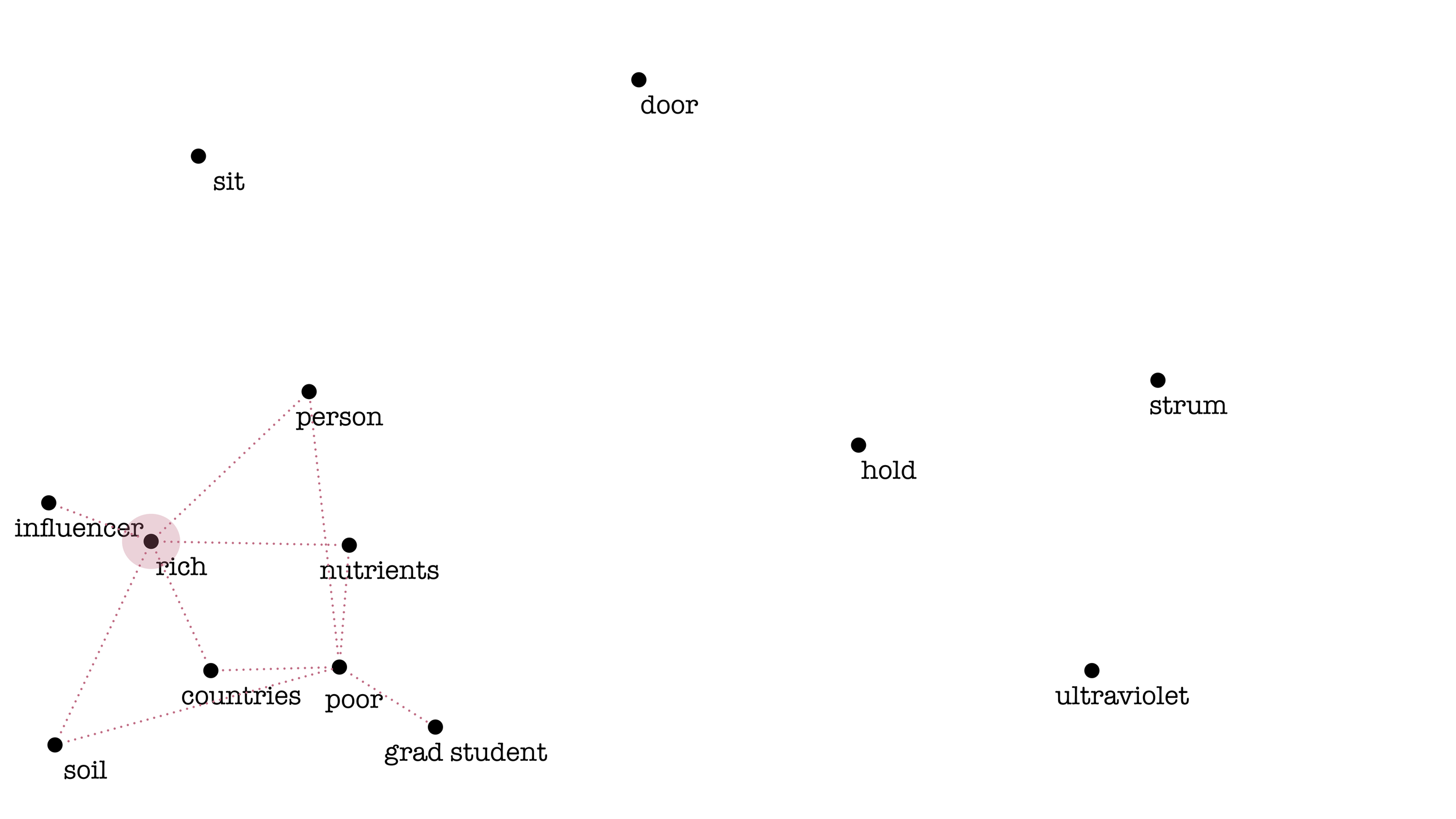

sit
person
influencer
rich
nutrients
countries
poor
grad student
soil

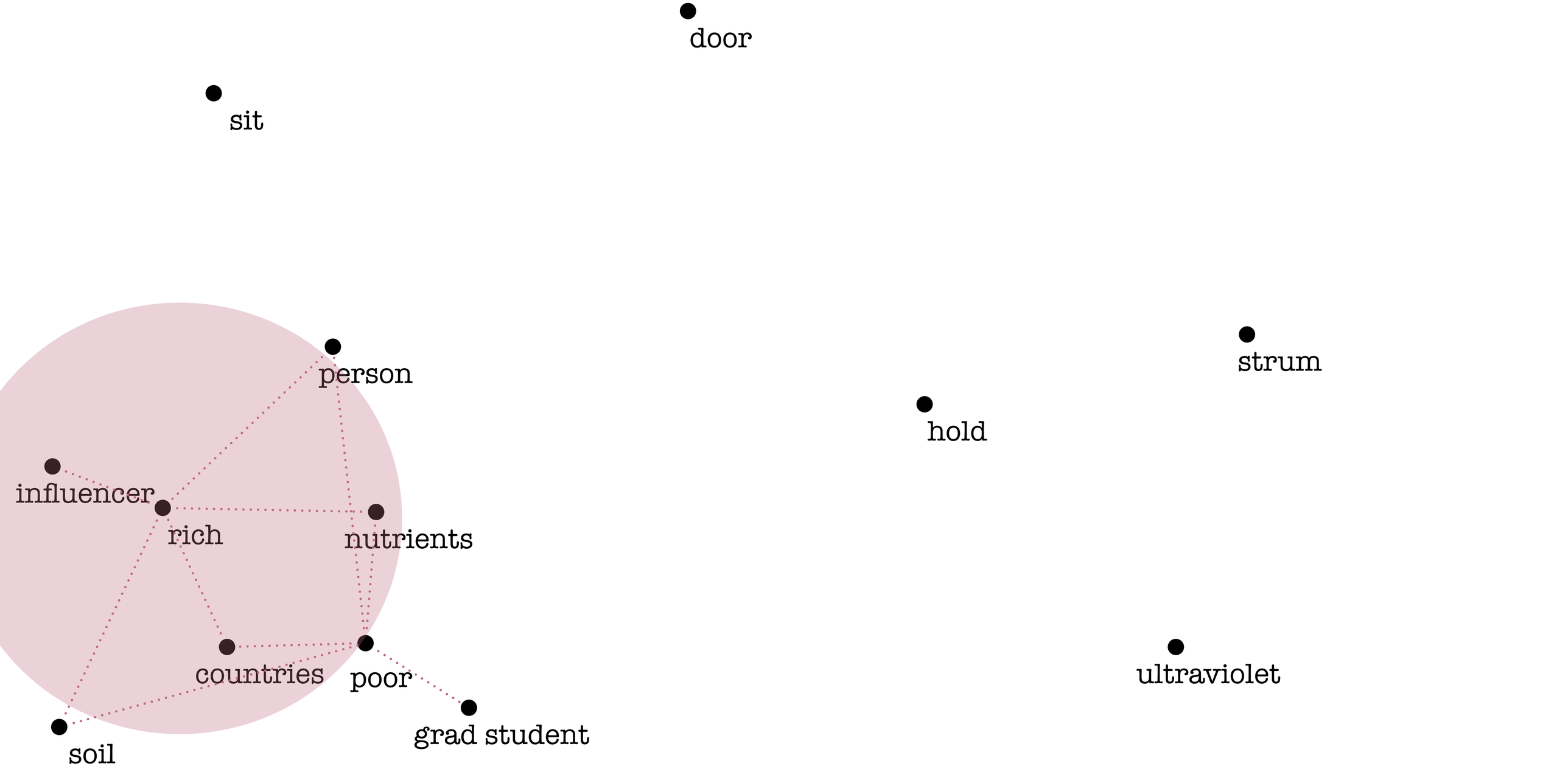

ultraviolet


strum


sit
door
person
influencer
rich
nutrients
countries
poor
grad student
soil
strum
ultraviolet

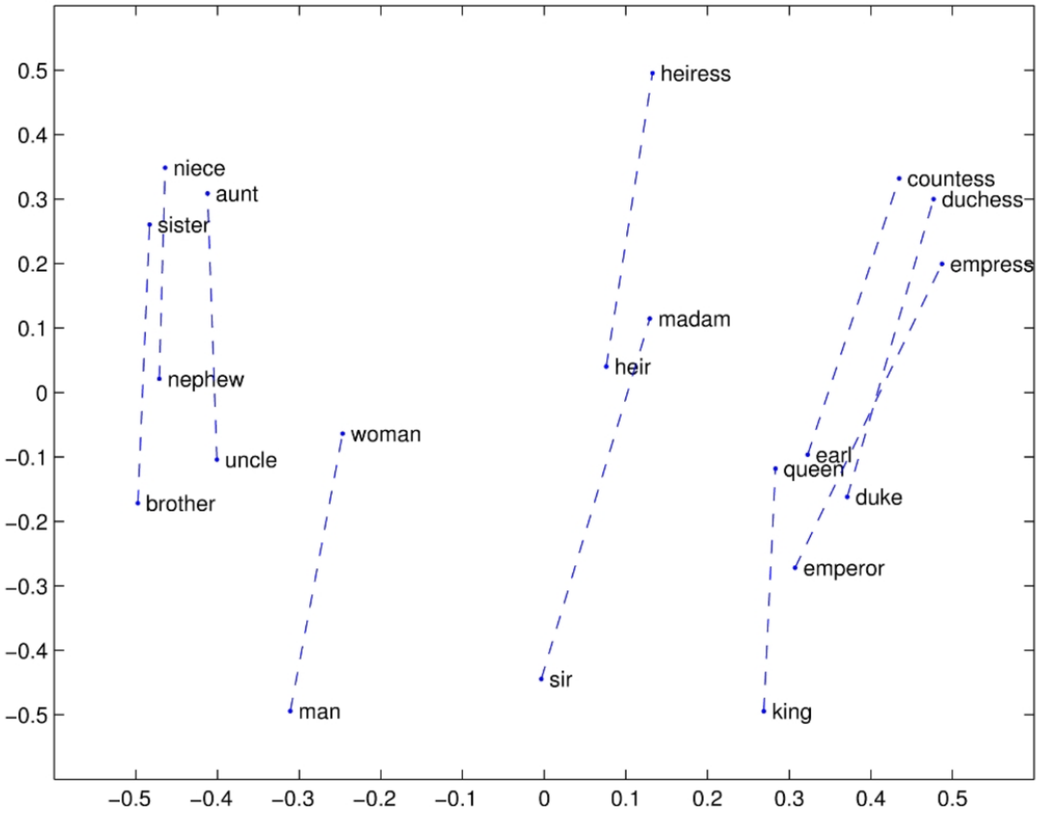

door
hold
sit
strum
person
influencer
rich
nutrients
countries
poor
ultraviolet
grad student
soil


door
sit
person
influencer
rich
nutrients
countries
poor
grad student
soil
strum
hold
ultraviolet


door
sit
person
strum
hold
influencer
rich
nutrients
countries
poor
ultraviolet
grad student
soil

Semantic properties on embeddings

Caveats: only seems to work for frequent words, small distances and certain relations, like relating countries to capitals, or parts of speech. [Linzen 2016, Gladkova et al. 2016, Ethayarajh et al. 2019a]

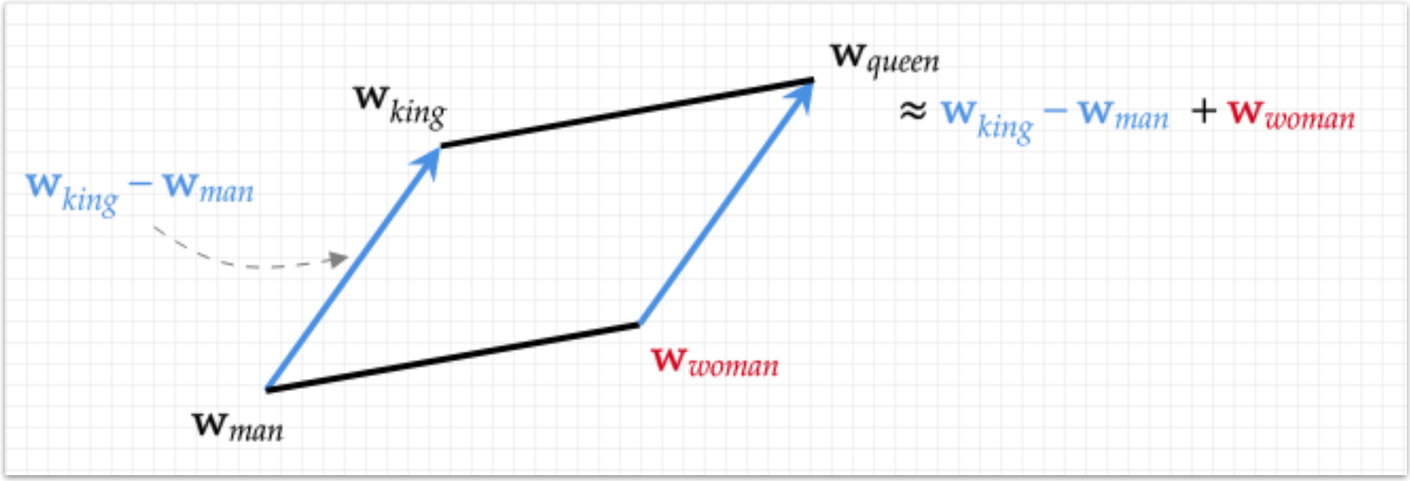


Tal Linzen. 2016. Issues in evaluating semantic spaces using word analogies. In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP*, pages 13–18, Berlin, Germany. Association for Computational Linguistics.

Naive applications

And possible pitfalls

- **defense** + **democrat** - **republican**
  - **democrat** and **republican** are likely to be nearest neighbors
  - **defense** + **democrat** - **republican** = **defense** + ϵ



Naive applications

And possible pitfalls

- **democrat - pathos** 🎭 **+ logos** ⚙️ **≈ republican**
  - republican is the more logical ⚙️ version of democrat?

- **democrat - pathos** 🎭 **+ logos** ⚙️ **= democrat + ϵ ≈ republican**

- **democrat + pathos** 🎭 **- logos** ⚙️ **= democrat - ϵ ≈ republican**
  - republican is the more emotional 🎭 version of democrat?

Distributed representations of words and phrases and their compositionality 42655 2013
T Mikolov, I Sutskever, K Chen, GS Corrado, J Dean
Neural information processing systems

Glove: Global vectors for word representation 39354 2014
J Pennington, R Socher, CD Manning
Proceedings of the 2014 conference on empirical methods in natural language …

Issues in evaluating semantic spaces using word analogies 178 2016
T Linzen
Proceedings of the First Workshop on Evaluating Vector Space Representations …

Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't. 260 2016
A Gladkova, A Drozd, S Matsuoka
Proceedings of the NAACL Student Research Workshop, 8-15

Caveats: only seems to work for frequent words, small distances and certain relations, like relating countries to capitals, or parts of speech. [Linzen 2016, Gladkova et al. 2016, Ethayarajh et al. 2019a]

Conclusion

- Best practices evolve quickly
- Computer scientists are well positioned to keep updated on these things
- Researchers aren’t incentivized to be explicit about shortcomings

- Our community is presenting models with certain promises
- Some parts of our community should help responsibly contextualize those

Conclusion

- Best practices evolve quickly
- Computer scientists are well positioned to keep updated on these things
- Researchers aren't incentivized to be explicit about shortcomings

- Our community is presenting **models** with certain promises
- Some parts of our community should help responsibly contextualize those

could be you!

Isn’t it easier to use an LLM?

Replicability

- Should be possible to understand what was done, and do the same experiment and get the same finding independently
- Why?
  - Detect errors and fraud
  - Assess how and in what ways findings are robust (or not)
- Leads to new scientific findings
- Provides basis for confidence in findings

http://arthurspirling.org/documents/BarriePalmerSpirling_TrustMeBro.pdf

Replicability with LLMs

- Using LLM for data labeling task
  - GPT-4 (OpenAI), Gemini 1.0/1.5 (Google), Llama-2 (Meta)
- Ran a data labeling task on each once a month
- High variance in responses, even when setting temperature
- Sometimes, backend changes resulted in inability to run same models again

http://arthurspirling.org/documents/BarriePalmerSpirling_TrustMeBro.pdf

1. **Take replication seriously, impose standards on ourselves and others.** Our most basic call is that researchers and their institutions—like journals—should be aware of the replication problems we discuss above. This issue is unlikely to 'go away' (we think it will get worse), and needs urgent action. Readers, referees and editors might consider down-weighting the contribution of papers that rely on routines that are unlikely to replicate—as they do currently for non-LM work.

2. **Consider open models that allow off-line versioning.** We found that, uniquely, our Llama implementation was replicable to a high standard *if* that standard is low variance. That is, if the goal is something approaching the Deterministic 'code and data' replication vision above, then local, versioned models are the way to go. These may not deliver top of the line performance (e.g. accuracy) but should be checked

Barrie, Christopher, Alexis Palmer, and Arthur Spirling. "Replication for Language Models Problems, Principles, and Best Practice for Political Science." *URL: https://arthurspirling. org/documents/BarriePalmerSpirling TrustMeBro. pdf* (2024)

How to Train Your Stochastic Parrot: Large Language Models for Political Texts*

Joseph T. Ornstein† Elise N. Blasingame‡ Jake S. Truscott§

July 23, 2024

Abstract

Large language models pretrained on massive corpora of text from the Internet have transformed the way that computer scientists approach natural language processing over the past five years. But such models have yet to see widespread adoption in the social sciences, partly due to their novelty and upfront costs. In this paper, we demonstrate how few-shot prompts to large language models can be effectively applied to a wide range of text-as-data tasks in political science—including sentiment analysis, document scaling, and topic modeling. In a series of pre-registered analyses, this approach outperforms conventional supervised learning methods without the need for extensive data pre-processing or large sets of labeled training data. And performance is comparable to expert and crowd-coding methods at a fraction of the cost. We propose a set of best practices for adapting these models to social science measurement tasks, and develop an open-source software package for researchers.

Ornstein JT, Blasingame EN, Truscott JS. How to train your stochastic parrot: large language models for political texts. Political Science Research and Methods. Published online 2025:1-18. doi:10.1017/psrm.2024.64

Putting all this together, we advise researchers to be cautious applying LLMs to tasks where a smart parrot spewing falsehoods, conspiracy theories, and hate speech would prove harmful.

Ornstein JT, Blasingame EN, Truscott JS. How to train your stochastic parrot: large language models for political texts. Political Science Research and Methods. Published online 2025:1-18. doi:10.1017/psrm.2024.64

But that's also kind of sad.

- These models undeniably have better performance, by many metrics
- Are there ways to use them that are valid for social science?

But that's also kind of sad.

- Social science research imposes different and interesting constraints on NLP algorithms
- These require additional work on and around the tools we present to social scientists
  - These are questions about NLP *models* that help *others* answer questions about people and their interactions

But that's also kind of sad.

- Social science research imposes different and interesting constraints on NLP algorithms
- These require additional work on and around the tools we present to social scientists
  - These are questions about NLP ***models*** that help *others* answer questions about people and their interactions

could be you!

Thank you!