

A background image showing a group of people in a meeting or collaborative work environment. A man on the right is gesturing with his hands while speaking to a group of people. The image is overlaid with a dark blue semi-transparent filter.

USE OF NATURAL LANGUAGE PROCESSING & KNOWLEDGE GRAPHS FOR ENHANCED DATA CONNECTIONS

Bilal Assaad

The Data Incubator
Data Science Fellowship
Winter Cohort
February 17, 2022

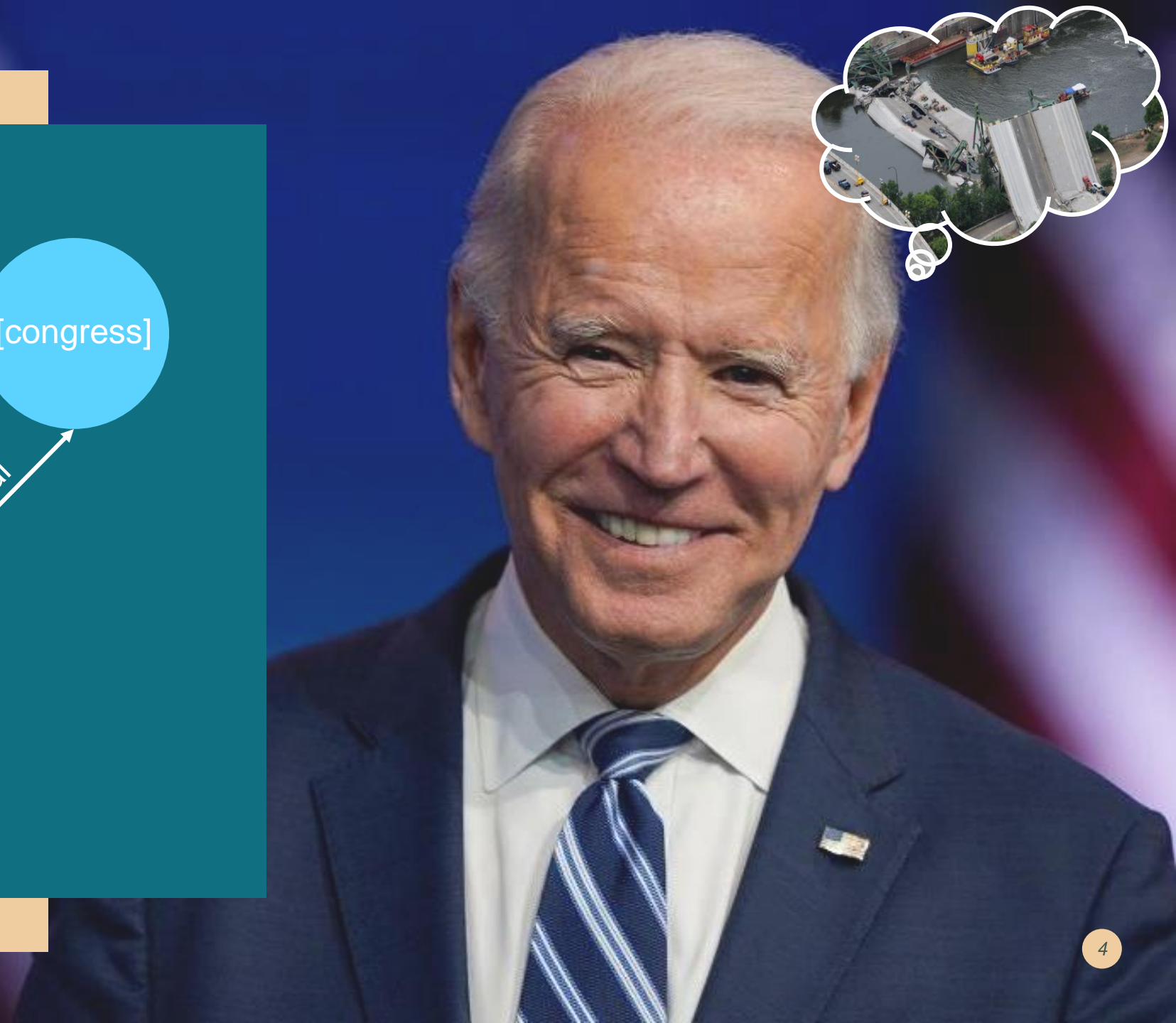
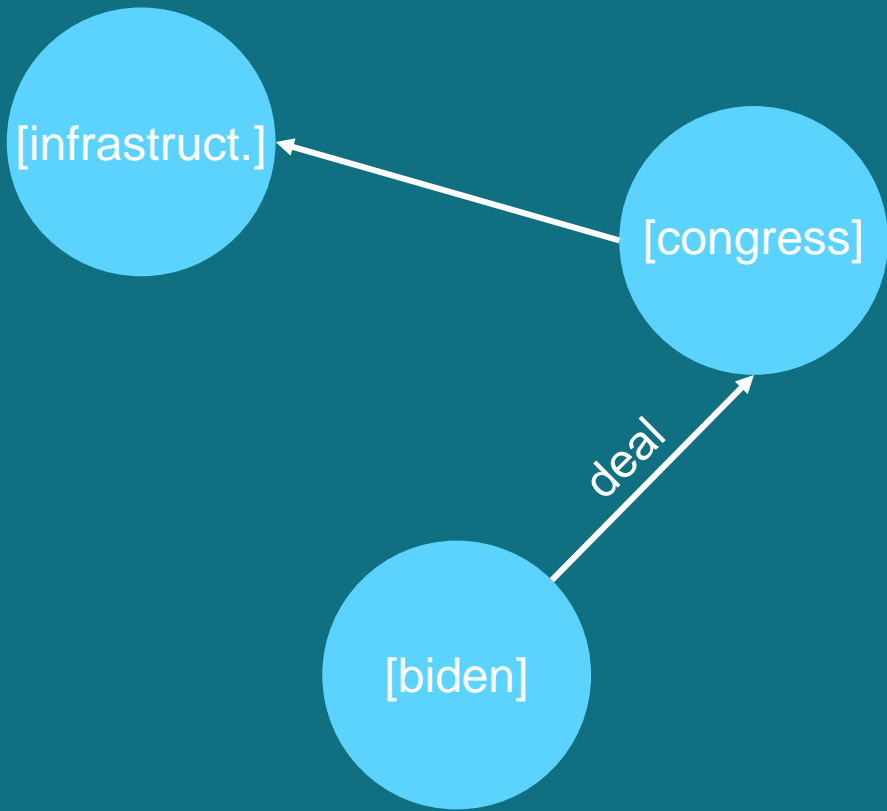


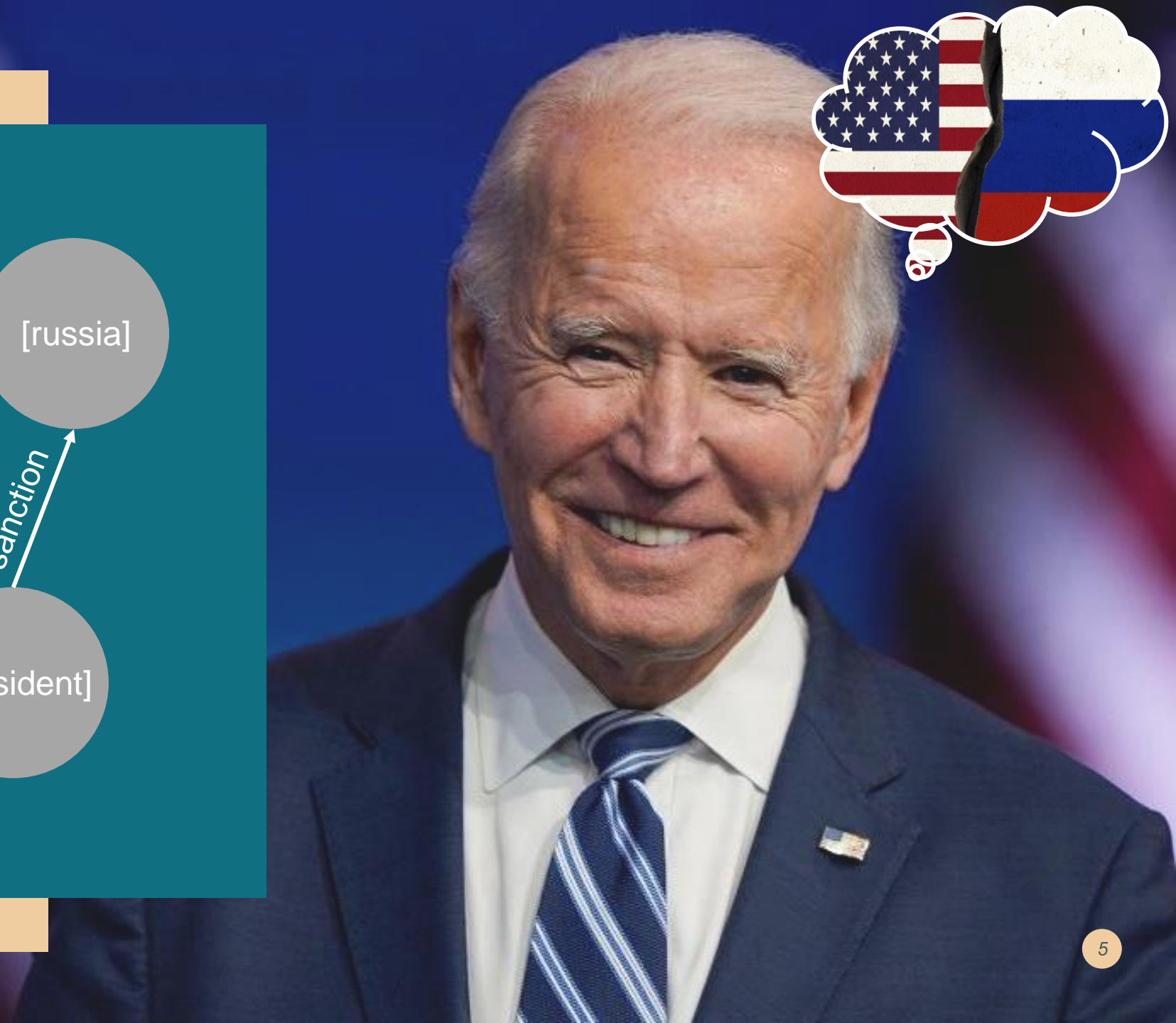
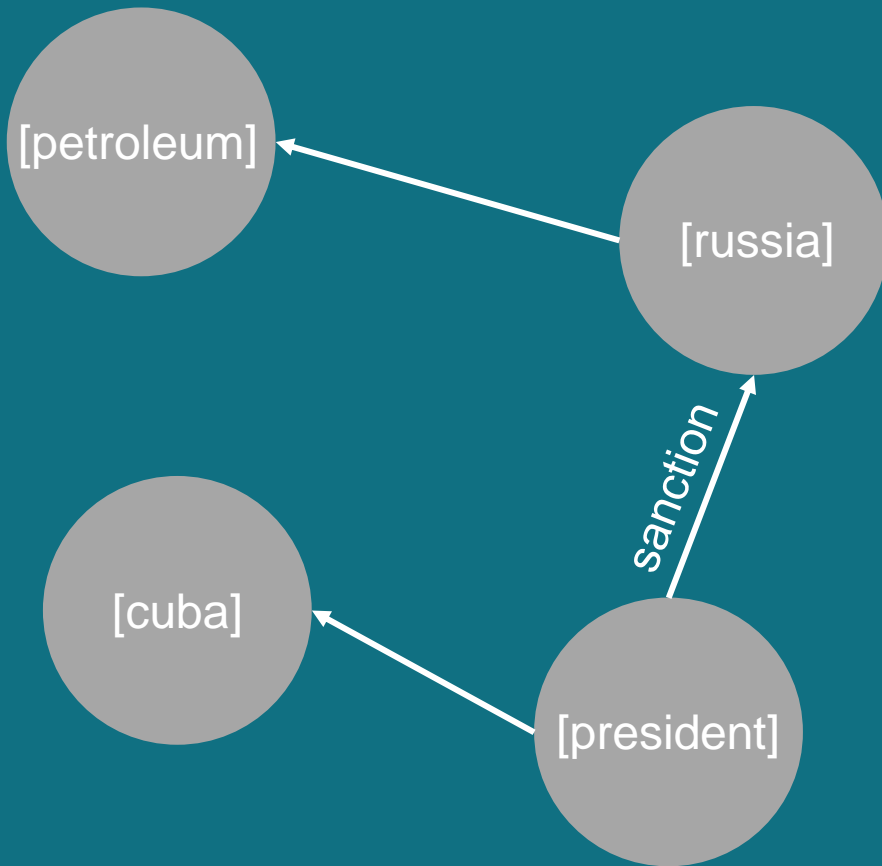
THE 'BIG' IDEA

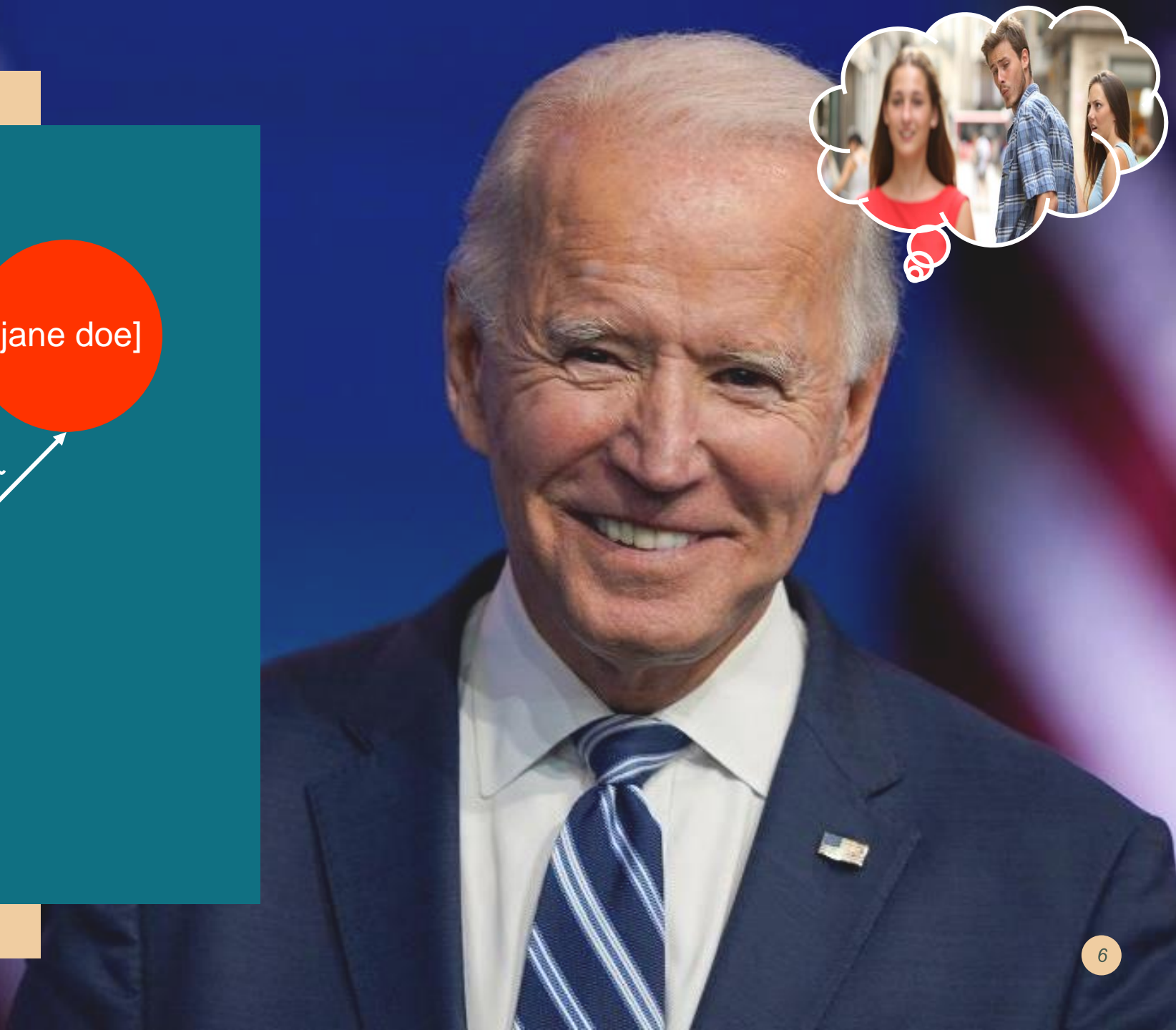
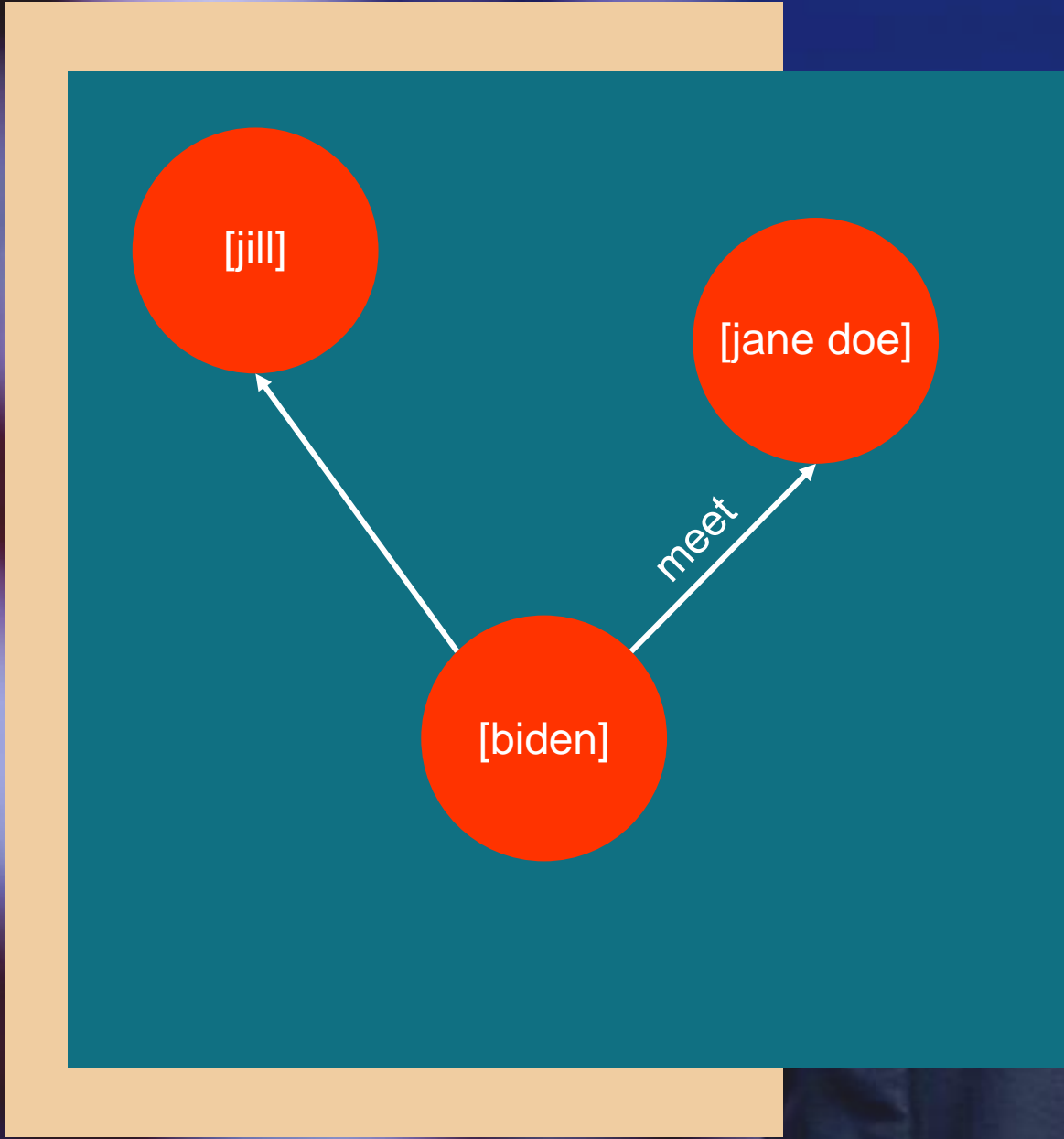
Make use of **Natural Language Processing (NLP)** in conjunction with **Knowledge Graphs (KG)** to obtain better **topic extraction, identification and visual relations** when conducting queries of published media contents in written form.

**EXAMPLE:
JOSEPH BIDEN #46**

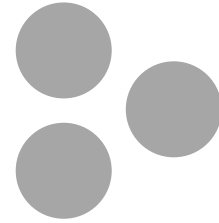
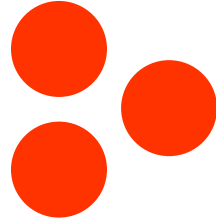
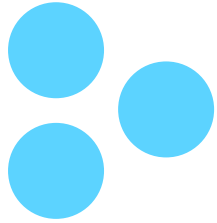








EXTRACTION OF TOPICS AND “BUCKETING”



**Celebrity
Gossip**



**Domestic
Policy**



**International
Policy**

COMPLAINTS ABOUT PRESENT METHODS



Users

Strategy teams

Marketing teams

Content/media managers



Present shortcomings

Generalized results

Obvious

ENHANCED SOLUTION



✓ Tool combination

NLP
+ KG
= NLPKG

✓ Address issues such as

Topic Extraction
Relationships

✓ Derive better

“What(s)”
“Why(s)”

MACHINE MODELING



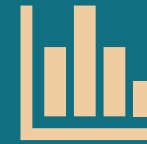
Data Ingestion & Text Preprocessing

- ~5K “docs”
- Each “doc” may contain ~20 to 2,000 words
- Remove miscellaneous text formatting



Train Model Using NLP

- Tokenization
- Lemmatization
- Stemming
- Named-Entity Recognition (NER)
- Relationships



Visualization & Knowledge Graph

- KG Nodes w/ Edges
- Word Frequency Vis.

STATIC DEMO

**MAKING SENSE OF
ACTUAL 'CONTENT' DATA**

TOPIC MODELING

Technique allows scanning of documents in order to cluster them.

SAMPLE PUBLISHED CONTENT

INPUT

WASHINGTON (AP) – Democratic leaders in Congress argue that impeaching President Donald Trump is a political mistake as the 2020 election nears.

Most of the candidates in the 2020 White House race seem to agree, for now.

Fewer than one-third of the 23 Democrats vying for the nomination are out with calls to start the impeachment process, citing evidence in special counsel Robert Mueller’s report that they believe shows Trump obstructed justice.

Most others, including leading contenders Joe Biden and Bernie Sanders, have stopped short.

That reluctance from the party’s presidential candidates, even as more congressional Democrats start pushing their leaders toward impeachment, underscores the risky politics of investigating the president for “high crimes and misdemeanors.”

Impeachment matters deeply to the party’s base, but remains unpopular with most Americans.

Copyright © 2019 The Associated Press. All rights reserved. This material may not be published, broadcast, written or redistributed.

OUTPUT

Topic 2 (~98%)



UNSEEN CONTENT

INPUT

Unseen Text >>> 'Speaker Pelosi leaves the U.S. Capitol to have an immediate meeting with the President to keep pressing our allies on the issues related to North Korea and climate change.'

OUTPUT

Topic 1 (~47%) ←

north + korea + kim + japan + missile + abe + korean

A horizontal bar chart representing the word cloud for Topic 1. The bar is divided into seven segments, each corresponding to a word: 'north', 'korea', 'kim', 'japan', 'missile', 'abe', and 'korean'. Each segment has a blue portion on the left and a gray portion on the right, with the blue portion representing the word's weight in the topic.

Topic 2 (~33%) ←

impeach + climate + conway + mueller + sanders + investigation + mccain

A horizontal bar chart representing the word cloud for Topic 2. The bar is divided into seven segments: 'impeach', 'climate', 'conway', 'mueller', 'sanders', 'investigation', and 'mccain'. Each segment has a blue portion on the left and a gray portion on the right, with the blue portion representing the word's weight in the topic.

Topic 3 (~18%)

debate + poll + senator + warren + sanders + fundraise + harris

A horizontal bar chart representing the word cloud for Topic 3. The bar is divided into seven segments: 'debate', 'poll', 'senator', 'warren', 'sanders', 'fundraise', and 'harris'. Each segment has a blue portion on the left and a gray portion on the right, with the blue portion representing the word's weight in the topic.

Topic 4 (~1%)

abort + climate + obama + plan + health + women + hyde

A horizontal bar chart representing the word cloud for Topic 4. The bar is divided into seven segments: 'abort', 'climate', 'obama', 'plan', 'health', 'women', and 'hyde'. Each segment has a blue portion on the left and a gray portion on the right, with the blue portion representing the word's weight in the topic.

Topic 5 (~1%)

debate + blasio + poll + city + cancer + buttigieg + mayor

A horizontal bar chart representing the word cloud for Topic 5. The bar is divided into seven segments: 'debate', 'blasio', 'poll', 'city', 'cancer', 'buttigieg', and 'mayor'. Each segment has a blue portion on the left and a gray portion on the right, with the blue portion representing the word's weight in the topic.

KNOWLEDGE GRAPHS

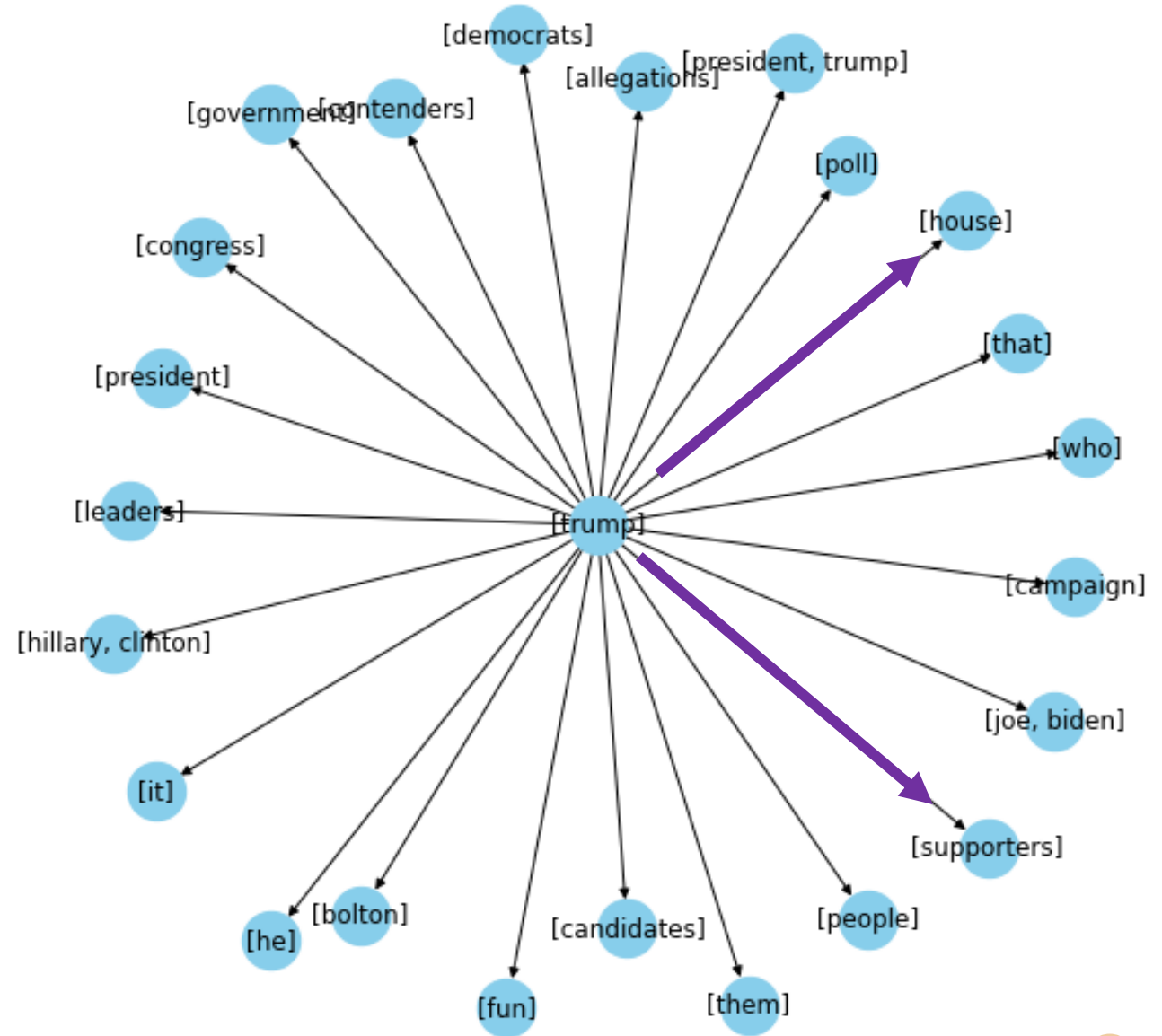
Semantic network represents real-world entities and their relationships.

CONTENT DEPTH LEVEL I

INPUT

Subject >>> 'trump'
Verb >>> 'meet'
Object >>> (multiple as determined by KG, see figure)
Depth >>> 1

OUTPUT

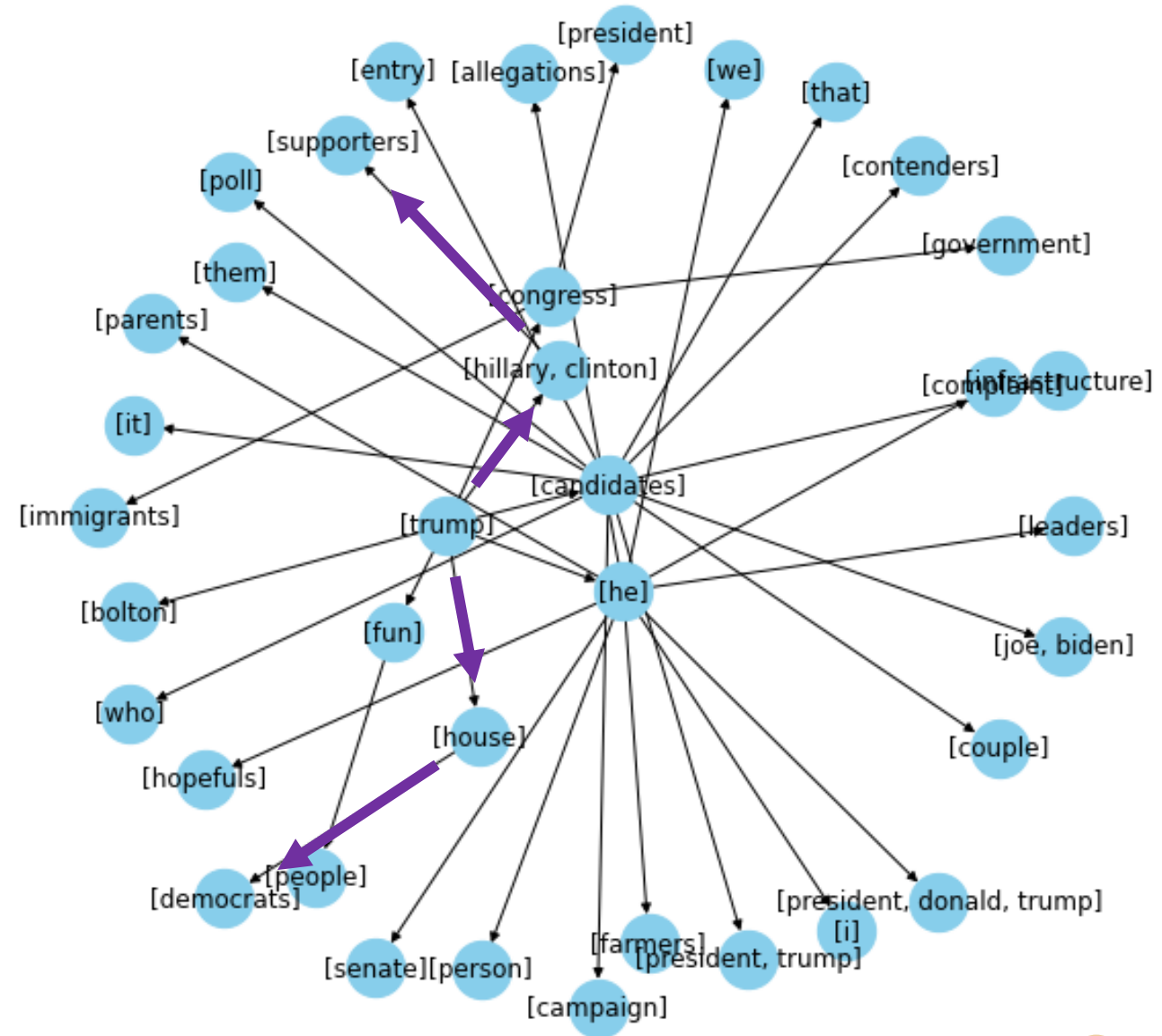


CONTENT DEPTH LEVEL 2

INPUT

Subject >>> 'trump'
Verb >>> 'meet'
Object >>> (multiple as determined by KG, see figure)
Depth >>> 2

OUTPUT



A background image showing three people (two men and one woman) smiling and looking at a screen, suggesting a collaborative work environment. The image is partially obscured by a teal overlay on the left side where the text is located.

SUMMARY

- ✓ **Increase Data Ingestion**
- ✓ **Better Topic Extraction**
- ✓ **Visual Data Representation**
- ✓ **Enhance Decision Making Process**

THANK YOU!



Bilal Assaad



bassaad17@gmail.com



[linkedin.com/in/bilalassaad](https://www.linkedin.com/in/bilalassaad)



github.com/bassaad17

