



# Computational Fact Checking from Knowledge Networks

G.L. Ciampaglia, P. Shiralkar, L.M. Rocha, J. Bollen, F. Menczer,  
A. Flammini [2015]



# Problem



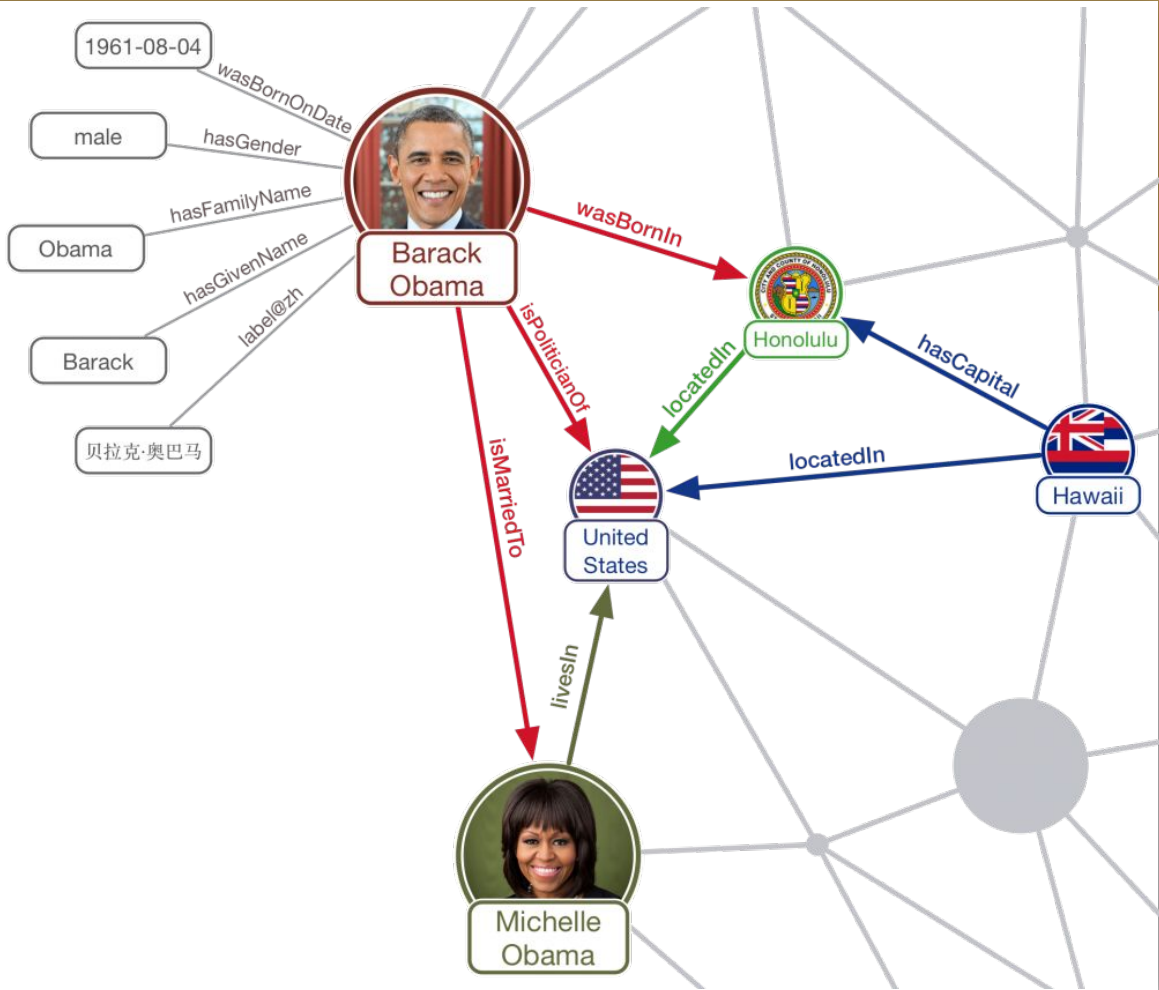
# Problem



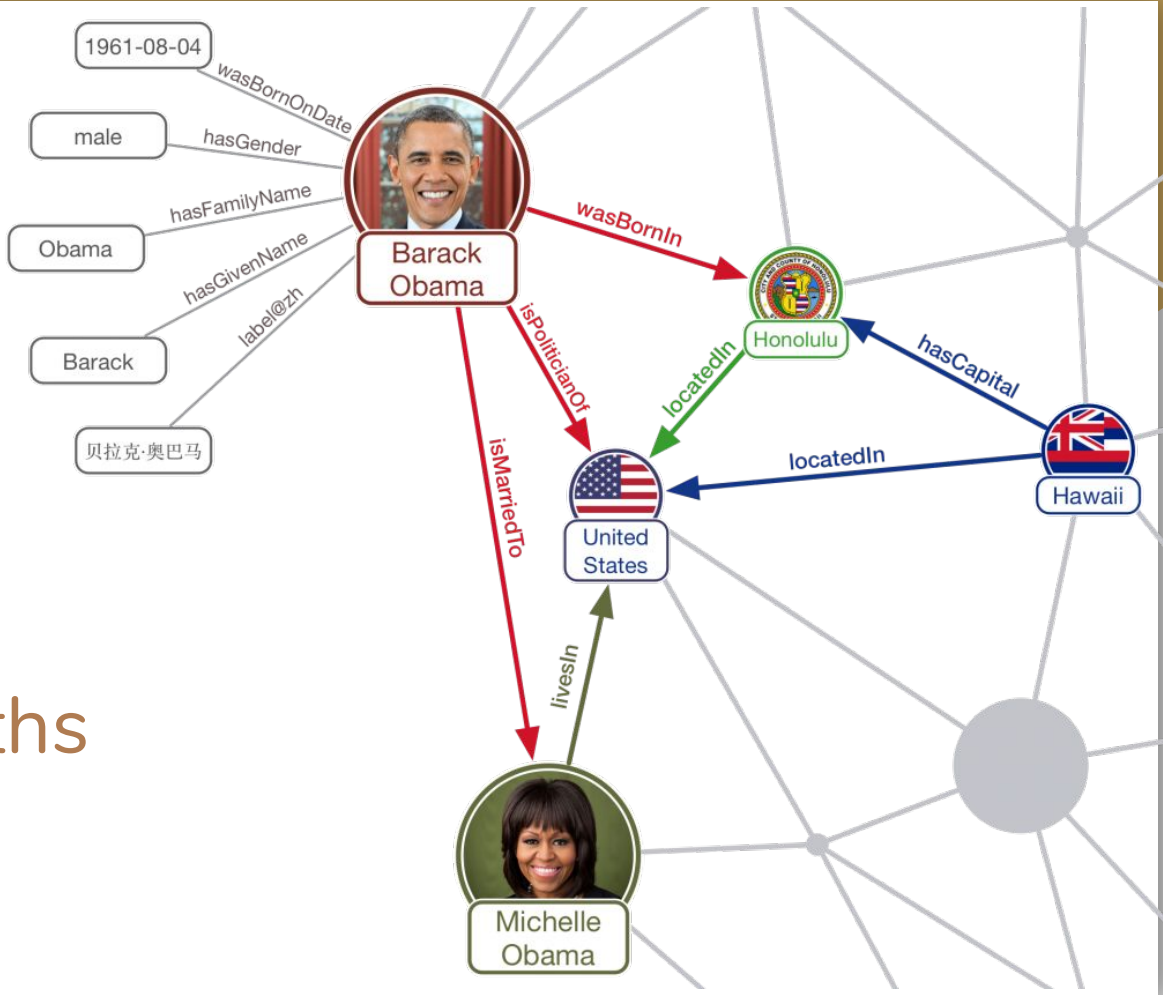
# What do we do?

1. Develop approaches to computationally check facts
2. Test & Validate
3. Scale
4. ???
5. Profit!

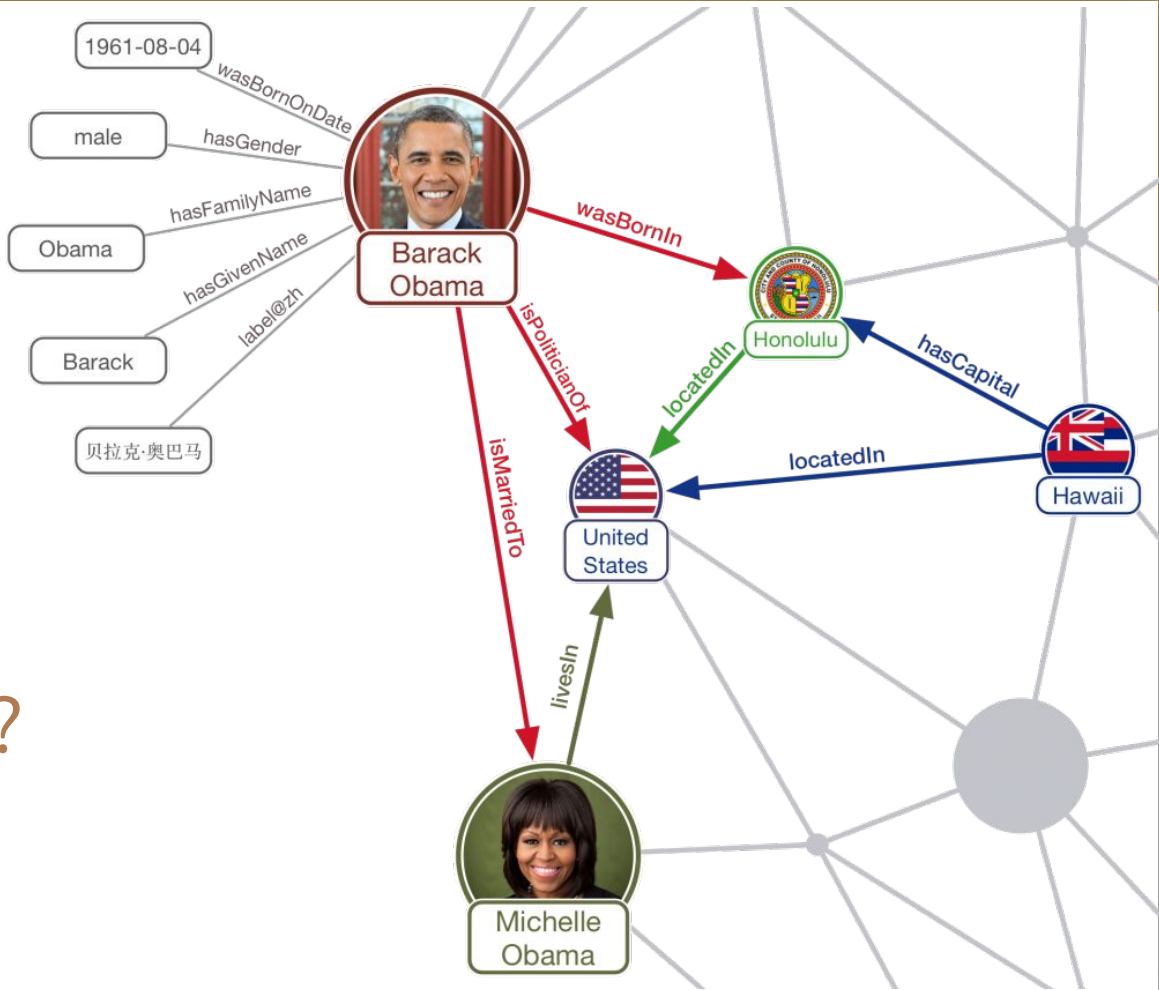
# The approach



# Weighted paths

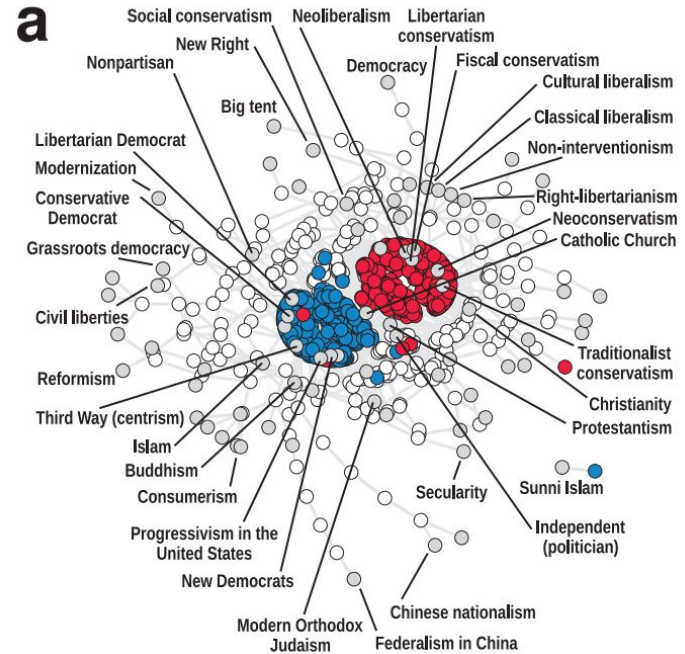


What is truth?



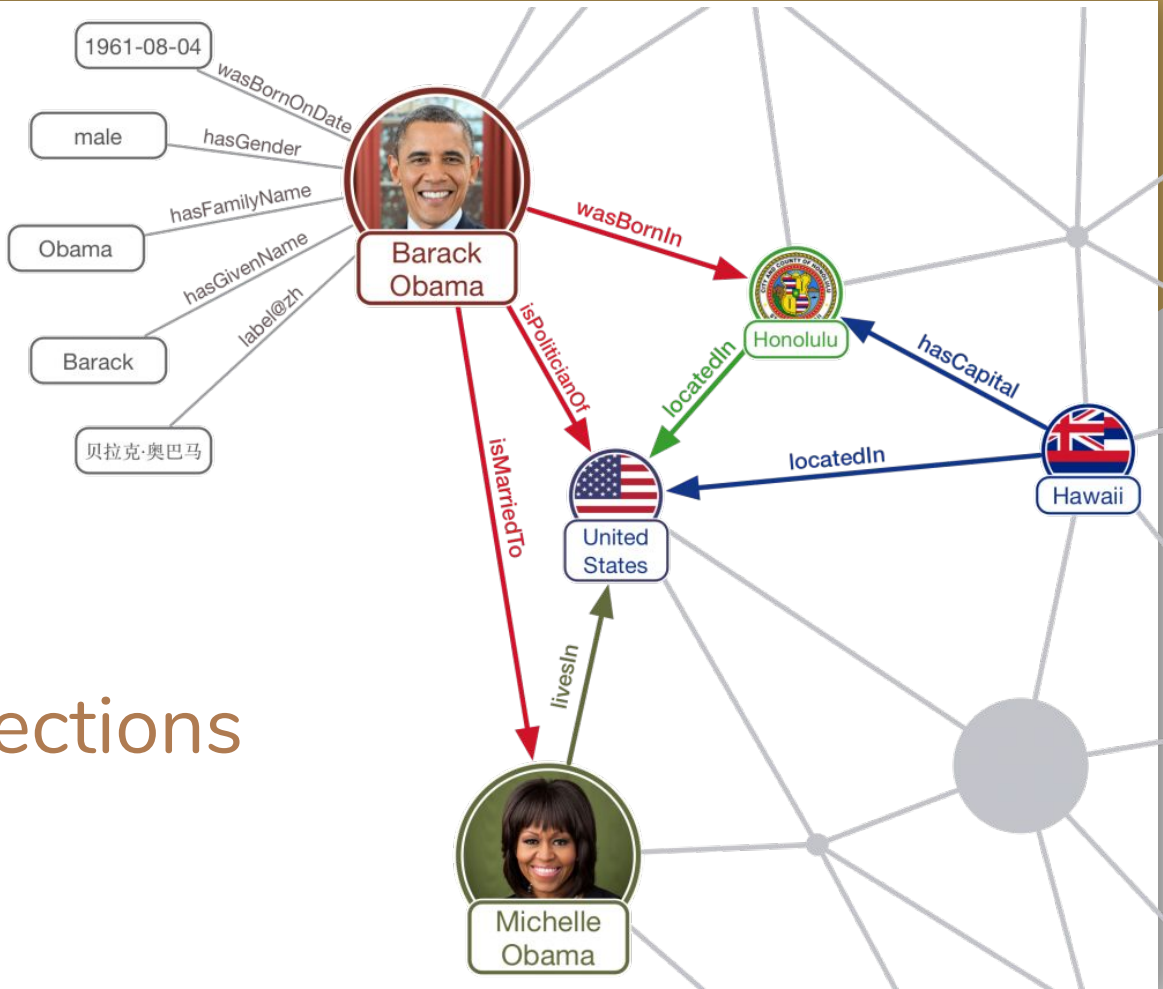
# Calibration test: Infer party affiliation of Congress members

vs. DW-NOMINATE





# Indirect Connections



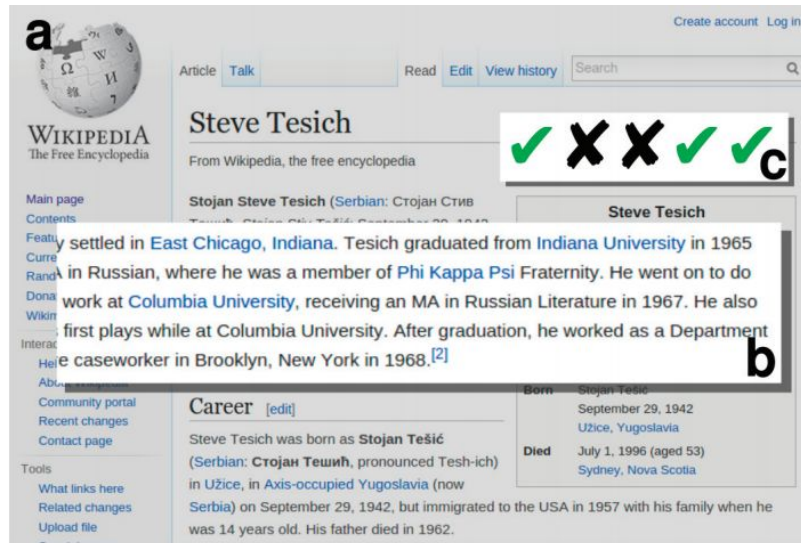
# How to validate?

## Annotated corpus

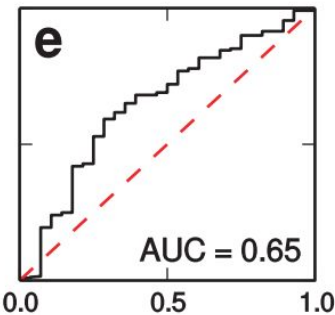
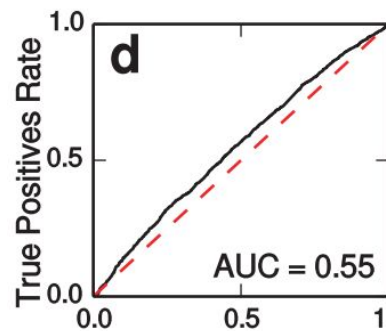
- Remove statements present in the WKG

## Results?

- Truth values positively correlated with average ratings by human evaluators
- Consistently higher support for true statements than false ones



The screenshot shows the Wikipedia article for Steve Tesich. The article text is partially visible, mentioning his birth in East Chicago, Indiana, and his education at Indiana University and Columbia University. A red box highlights a sentence: "settled in East Chicago, Indiana. Tesich graduated from Indiana University in 1965 in Russian, where he was a member of Phi Kappa Psi Fraternity. He went on to do work at Columbia University, receiving an MA in Russian Literature in 1967. He also first plays while at Columbia University. After graduation, he worked as a Department e caseworker in Brooklyn, New York in 1968.<sup>[2]</sup>". Above this text, there are five green checkmarks and two black X's, indicating human evaluations. A red box labeled 'b' highlights the 'Career' section, which states: "Steve Tesich was born as Stojan Tešić (Serbian: Стојан Тешић, pronounced Tesh-ich) in Uzice, in Axis-occupied Yugoslavia (now Serbia) on September 29, 1942, but immigrated to the USA in 1957 with his family when he was 14 years old. His father died in 1962." The 'Born' and 'Died' sections are also visible, providing dates and locations.



False Positives Rate

# Key Points

Simple shortest path computation maximizing information content can leverage an existing body of collective human knowledge to assess truth.

This approach utilizes topology of knowledge graphs rather than explicit content of statements (and is more performant for it).

- Indicates value of indirect paths for the task

# Questions

1. Say a news aggregator (e.g. Google News) implements this as a scoring system and favors high scores. Who else do you think will most quickly adopt this technology? (Hint: Who stands to gain the most?)

# Questions

1. Say a news aggregator (e.g. Google News) implements this as a scoring system and favors high scores. Who else do you think will most quickly adopt this technology? (Hint: Who stands to gain the most?)
2. Is this approach biased toward mainstream information? What about radical but true (e.g. revelations from whistleblowers, secrets, scandals)?

# Questions

1. Say a news aggregator (e.g. Google News) implements this as a scoring system and favors high scores. Who else do you think will most quickly adopt this technology? (Hint: Who stands to gain the most?)
2. Is this approach biased toward mainstream information? What about radical but true (e.g. revelations from whistleblowers, secrets, scandals)?
3. How could an adversary attack this system?