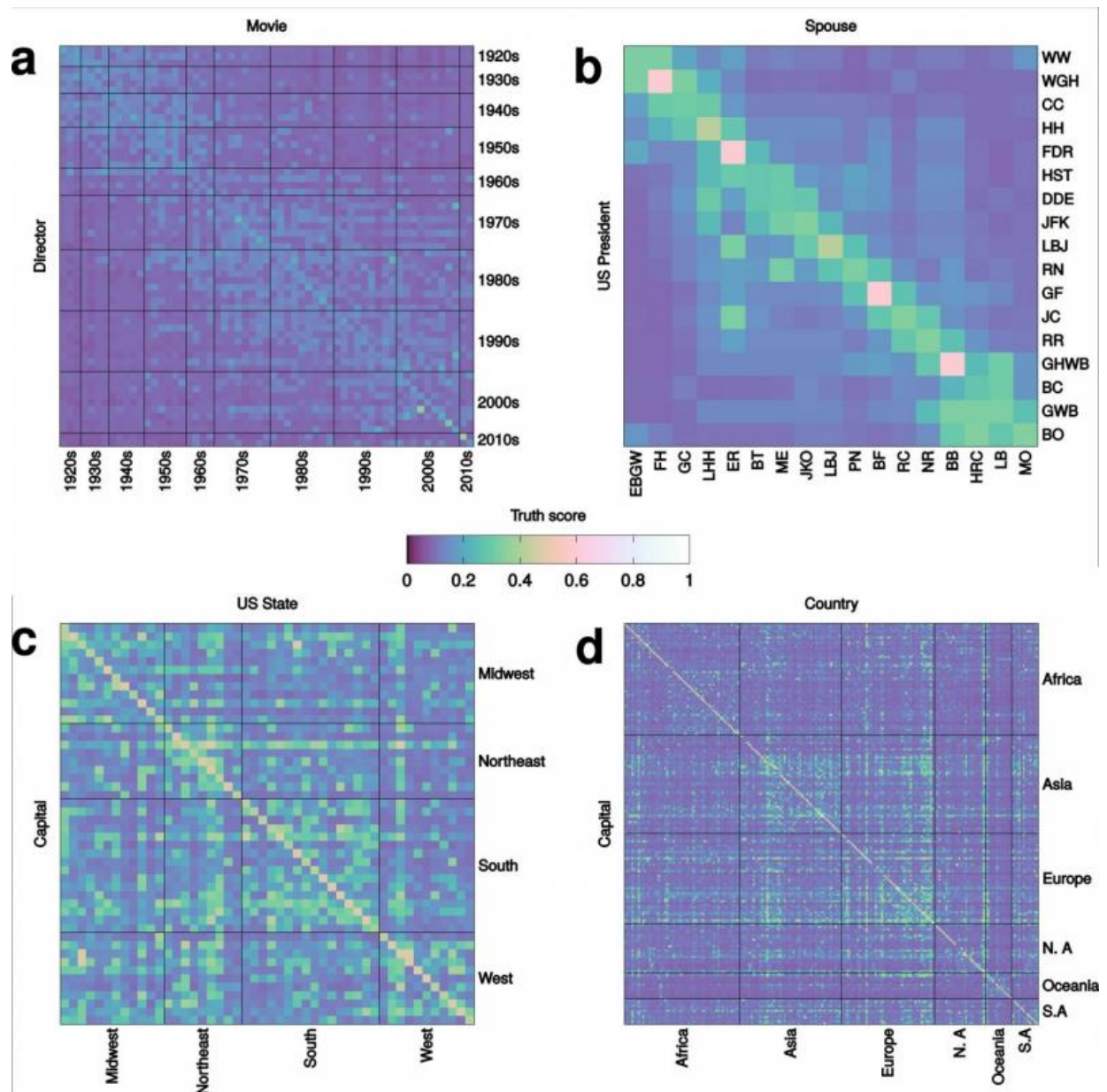# Scientists create computational algorithm for fact-checking

June 17 2015

These charts map the 'truth scores; of statements related to geography, history and entertainment. Correct statements appear along the diagonal, with color intensity indicating the strength of the score. Credit: Giovanni Ciampaglia

Network scientists at Indiana University have developed a new computational method that can leverage any body of knowledge to aid in the complex human task of fact-checking.

In the first use of this method, IU scientists created a simple computational fact-checker that assigns "truth scores" to statements concerning history, geography and entertainment, as well as random statements drawn from the text of Wikipedia, the well-known online encyclopedia.

In multiple experiments, the automated system consistently matched the assessment of human fact-checkers in terms of their certitude about the accuracy of these statements.

The results of the study, "Computational Fact Checking From Knowledge Networks," are reported in today's issue of *PLOS ONE.*

"These results are encouraging and exciting," said Giovanni Luca Ciampaglia, a postdoctoral fellow at the Center for Complex Networks and Systems Research in the IU Bloomington School of Informatics and Computing, who led the study. "We live in an age of information overload, including abundant misinformation, unsubstantiated rumors and conspiracy theories whose volume threatens to overwhelm journalists and the public.

"Our experiments point to methods to abstract the vital and complex human task of fact-checking into a network analysis problem, which is

easy to solve computationally."

The team selected Wikipedia as the information source for their experiment due to its breadth and open nature. Although Wikipedia is not 100 percent accurate, previous studies estimate the [online encyclopedia](link) is nearly as reliable as traditional encyclopedias, but also covers many more subjects.

Using factual information from infoboxes on the site, IU scientists built a "knowledge graph" with 3 million concepts and 23 million links between them. A link between two concepts in the graph can be read as a simple factual statement, such as "Socrates is a person" or "Paris is the capital of France."

In what the IU scientists describe as an "automatic game of trivia," the team applied their algorithm to answer simple questions related to geography, history and entertainment, including statements that matched states or nations with their capitals, presidents with their spouses and Oscar-winning film directors with the movie for which they won the Best Picture awards, with the majority of tests returning highly accurate truth scores.

Lastly, the scientists used the algorithm to fact-check excerpts from the main text of Wikipedia, which were previously labeled by human fact-checkers as true or false, and found a positive correlation between the truth scores produced by the algorithm and the answers provided by the fact-checkers.

Significantly, the IU team found their [computational method](link) could even assess the truthfulness of statements about information not directly contained in the infoboxes. For example, the fact that Steve Tesich —-the Serbian-American screenwriter of the classic Hoosier film "Breaking Away"—graduated from IU, despite the information not being

specifically addressed in the infobox about him.

"The measurement of the truthfulness of statements appears to rely strongly on indirect connections, or 'paths,' between concepts," Ciampaglia said. "If we prevented our fact-checker from traversing multiple nodes on the graph, it performed poorly since it could not discover relevant indirect connections. But because it's free to explore beyond the information provided in one infobox, our method leverages the power of the full knowledge graph."

Although the experiments were conducted using Wikipedia, the IU team's method does not assume any particular source of knowledge. The scientists aim to conduct additional experiments using knowledge graphs built from other sources of human knowledge, such as Freebase, the open-knowledge base built by Google, and note that multiple information sources could be used together to account for different belief systems.

"Misinformation endangers the public debate on a broad range of global societal issues," said Filippo Menczer, director of the Center for Complex Networks and Systems Research and a professor in the IU School of Informatics and Computing, who is a co-author on the study. "With increasing reliance on the Internet as a source of information, we need tools to deal with the misinformation that reaches us every day. Computational fact-checkers could become part of the solution to this problem."

The team added a significant amount of natural language processing research, and other work remains before these methods could be made available to the public as a software tool.

  More information: *PLOS ONE,*
dx.plos.org/10.1371/journal.pone.0128193

Provided by Indiana University