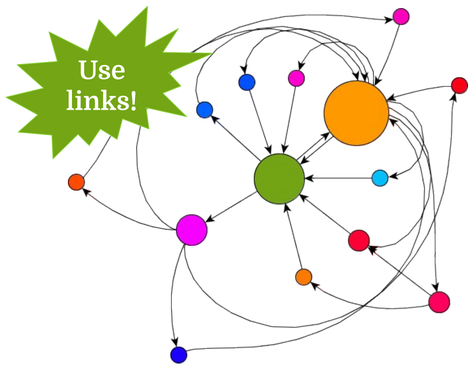
# Discovering Topical Context from Links in Wikipedia

**Cristian Consonni**[1], David Laniado[2], Alberto Montresor[1]

[1] DISI, University of Trento – [2] Eurecat - Centre Tecnològic of Catalunya

## Problem

**Can we find the context of a given topic from the graph of internal links in Wikipedia?**




## Data: WikiLinkGraphs

| lang | size | N | E |
|---|---|---|---|
| de | 5.7 | 3,588,883 | 59,535,864 |
| en | 17.0 | 13,685,337 | 163,380,007 |
| es | 3.0 | 3,034,113 | 38,348,163 |
| fr | 4.8 | 3,443,206 | 57,823,305 |
| it | 3.1 | 2,117,022 | 37,814,105 |
| nl | 2.0 | 2,626,527 | 25,834,057 |
| pl | 2.3 | 1,684,606 | 25,901,789 |
| ru | 3.2 | 3,360,531 | 37,394,229 |
| sv | 2.0 | 6,131,736 | 52,426,633 |

**Statistics about the WikiLinkGraphs dataset: Wikipedia language edition (*lang*), size in GB of the data (*size*), number of nodes (*N*) and edges (*E*) of the latest graph snapshot (2018-03-01).**

Consonni, C, Laniado D., Montresor A.
***WikiLinkGraphs:***
***A complete, longitudinal and multi-language dataset of the Wikipedia link networks.***
To appear at ICWSM 2019.

**https://zenodo.org/record/2539424**

## Approach

### LoopRank

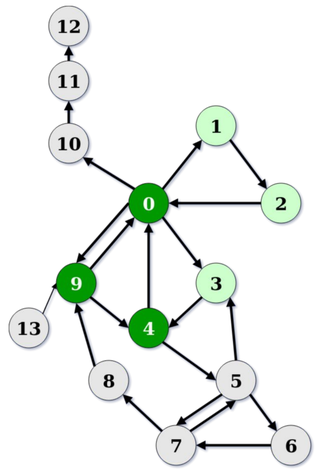
**Given a graph G and a reference node *r*, LoopRank assignes a score to each node depending on the number and the length of simple loops of maximum length *K* that go through the reference node *r*.**

$$LR_{r,K}(i) = \sum_{n=2}^{K} \sigma(n) \cdot \ell_{r,n}(i) = \sum_{n=2}^{K} \frac{\ell_{r,n}(i)}{n} \qquad (1)$$

```
# of cycles found: 4

--> cycle: 0-1-2-0
--> cycle:
0-3<<<-4-0
--> cycle: 0-9-0
--> cycle: 0-9-4-0


score(0): 1.500000
score(1): 0.333333
score(2): 0.333333
score(3): 0.333333
score(4): 0.666667
score(9): 0.833333
```



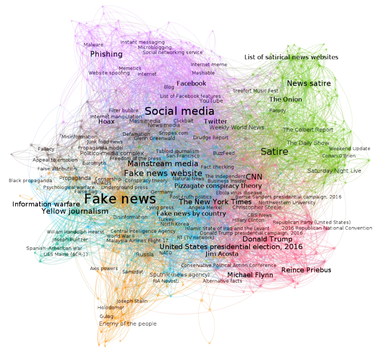

### Sketch of the algorithm

1. BFS from $r$ on the graph $G(V,E)$ to calculate $d(r)$;
2. Discard immediately all the nodes at a distance $d(r)>K-1$;
3. BFS from $r$ on the transposed network $G^T(V,E)$ to calculate $d^T(r)$;
4. Discard all the nodes for which $d(r)+d^T(r)>K$;
5. Enumerate all loops starting from $r$ using Johnson's algorithm;
6. Assign the LoopRank score as per Equation 1.



## Case Study: FAKE NEWS

**Fake news** or **junk news** or **pseudo-news** is a type of yellow journalism or propaganda that consists of deliberate disinformation or hoaxes spread via traditional print and broadcast news media or online social media.




**2017**

**2018**


**Graphs induced by the nodes with non-zero LoopRank score with reference node r = "Fake news" and K = 4 on English Wikipedia. over the snapshot of March 1st, 2018. Colors represents clusters calculated with the Louvain algorithm. The dimension of the nodes and their labels depends on the LoopRank score.**

| # | de | it | fr | it |
|---|---|---|---|---|
| 1 | Fake News | Fake news | Fake news | Fake news |
| 2 | Barack Obama | (2) CNN | (3) Donald Trump | (1) Disinformazione |
| 3 | (2) Tagesschau.de | (4) Facebook | Élection présidentielle française de 2017 | Post-verità |
| 4 | (3) Donald Trump | (3) United States presidential election, 2016 | (4) Facebook | (1) Bufala |
| 5 | (1) Desinformation | Social media | Ère post-vérité | Debunker |
| 6 | (3) Donald Trumps Präsidentschafts-wahlkampf 2015/16 | (1) Propaganda | Emmanuel Macron | (1) Manipolazione dell'informazione |
| 7 | (2) Der Freitag | (3) Donald Trump presidential campaign, 2016 | Guerre civile syrienne | Verifica dei fatti |
| 8 | (3) Präsidentschaftswahl in den Vereinigten Staaten 2016 | (2) The New York Times | (1) Désinformation | (1) Clickbait |
| 9 | Postfaktische Politik | Fake news website | (1) Rumeur | Spin doctor |
| 10 | Hillary Clinton | Pope Francis | (2) Conspiracy Watch | (2) Candido (rivista) |

**Top-10 articles with the highest LoopRank score computed from the page "Fake news" or equivalent in the given language, over the most recent snapshot of the WikiLinkGraphs dataset (2018-03), for German (de), English (en), French (fr), and Italian (it) Wikipedia.**

## Acknowledgements

This work has been supported by the European Union's Horizon 2020 research and innovation programme under the EU Engineroom project, with Grant Agreement n° 780643.