

Navigation, Networks, News, Wikipedia



Patrick Gildersleve
Oxford Internet Institute



Taha Yasseri
UCD



Renaud Lambiotte
Mathematical Institute



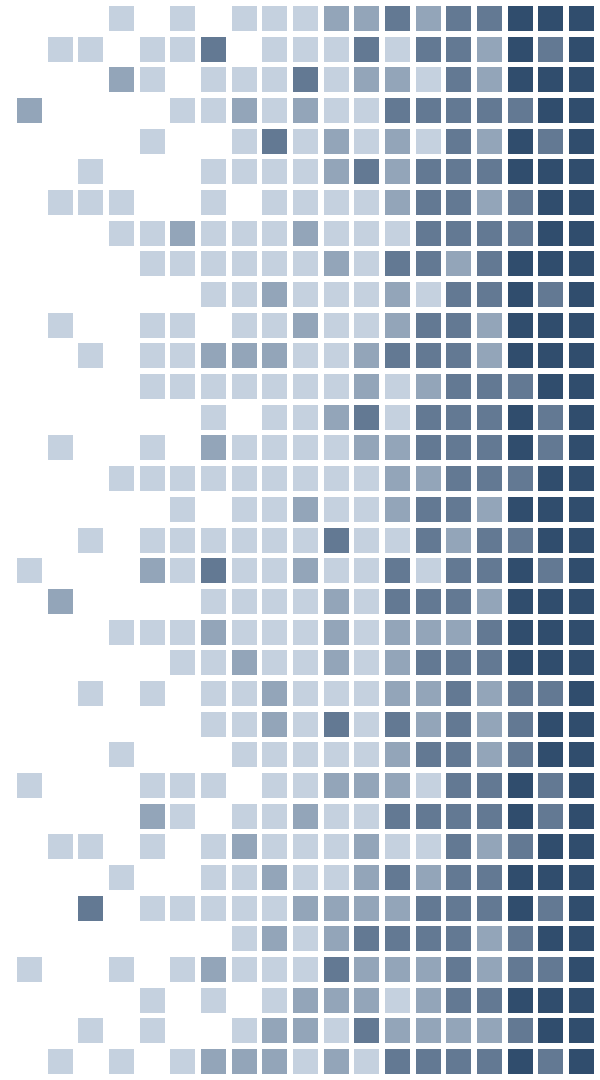
Outline

1. Navigation patterns on Wikipedia
2. News events and collective attention



1.

Navigation patterns on Wikipedia





How does the organisation of information on Wikipedia shape user traffic?

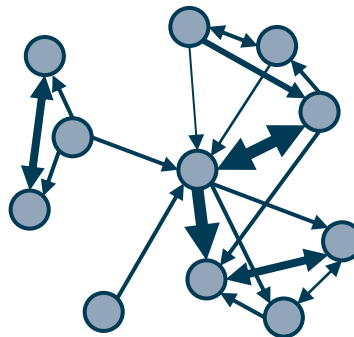
Gildersleve, P., & Yasseri, T. (2018). Inspiration, captivation, and misdirection: Emergent properties in networks of online navigation. *Complex Networks IX (CompleNet 2018)*.



Wikipedia Clickstream

- Month long counts for transitions on hyperlinks between all (English) Wikipedia articles
- Weighted, directed network of Wikipedia articles

	prev	curr	type	n
0	Economics_(Aristotle)	Aristotle	link	12
1	Greek_words_for_love	Aristotle	link	14
2	Pythias	Aristotle	link	12
3	Psychology	Aristotle	link	37
4	Aristotelian_theology	Aristotle	link	16
5	Peripatetic_school	Aristotle	link	89
6	Homer	Aristotle	link	57
7	Albertus_Magnus	Aristotle	link	19
8	Praxis_(process)	Aristotle	link	20
9	Astrophysics	Aristotle	link	12



Sources & Sinks, Bottlenecks & Distributors

- Volume of traffic through articles
- Sourcing coefficient:

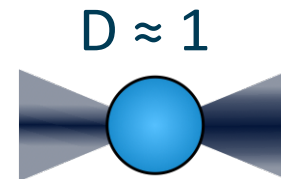
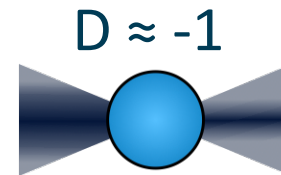
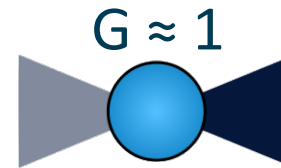
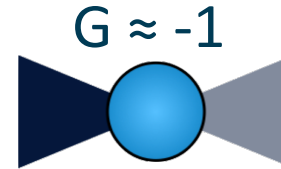
- $$G = \frac{S_{out} - S_{in}}{S_{out} + S_{in}}$$

- Diversity of traffic through articles
- Spreading coefficient:

- $$D = \bar{\sigma}_{out} - \bar{\sigma}_{in}$$

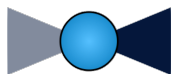
- $$\bar{\sigma} = -\frac{1}{\ln(k)} \sum_{i \in edges} \frac{w_i}{S} \ln\left(\frac{w_i}{S}\right)$$

$k = in/out\ degree, S = in/out\ strength,$
 $w = edge\ weight$



Navigational Phase Space - all articles

Sources



Sourcing Coefficient

1

0

-1

-1

0

1

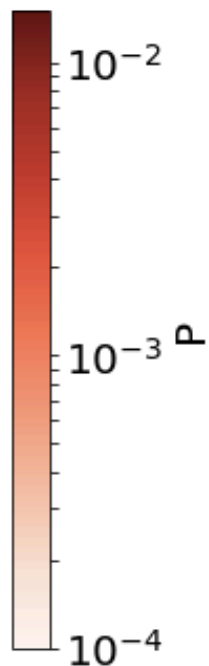
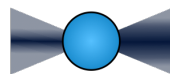
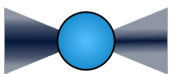
Spreading Coefficient

0

1

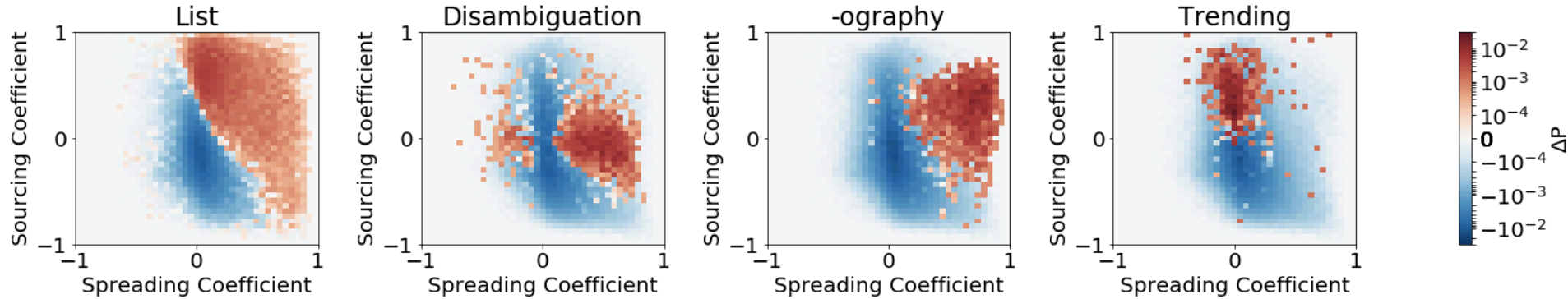
Distributors

Bottlenecks



Distinct Navigational Roles

Sourcing Coefficient (G) vs Spreading Coefficient (D) Difference Heatmaps



- Different sourcing / spreading behaviour
- Identify deficiencies & errors in the article network

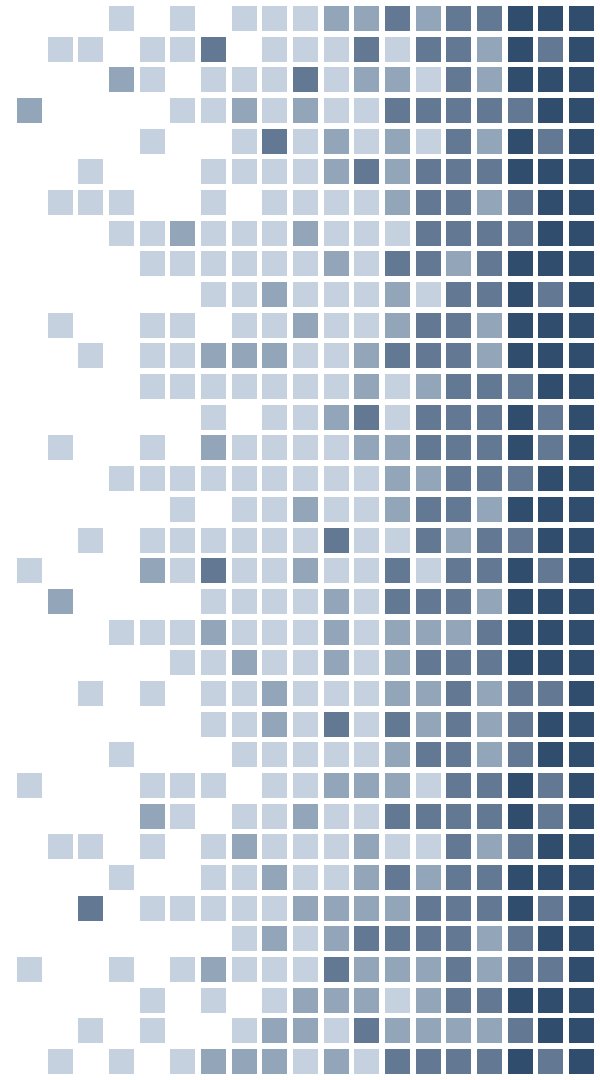
Conclusions

- Different navigation patterns for different article types
- Use sourcing/spreading behaviour to identify deficiencies & errors in the article network
- Demonstration of wider network science, using traffic to study the underlying structure



2.

News & Collective Attention





Katherine Maher ✓

@krmaher



As usual, I learn as much from the Wikipedia articles about breaking news events as I do from the news itself.

en.wikipedia.org/wiki/2020_Beir...

Various customs officials had sent letters to judges requesting a resolution to the issue of the confiscated cargo, proposing that the ammonium nitrate either be exported, given to the [Army](#), or sold to the private Lebanese Explosives Company.^{[c][13]} Letters had been sent on 27 June 2014, 5 December 2014, 6 May 2015, 20 May 2016, 13 October 2016, and 27 October 2017.^[13] One of the letters sent in 2016 noted that judges had not replied to previous requests, and "pleaded":^[13]

In view of the serious danger of keeping these goods in the hangar in unsuitable climatic conditions, we reaffirm our request to please request the marine agency to re-export these goods immediately to preserve the safety of the port and those working in it, or to look into agreeing to sell this amount

ATENÇÃO
SISTEMA DE
VIGILANCIA
EM OPERACAO

NOVO CONCEITO
DE SEGURANCA
COM FERRAMENTAS
DE PROTECCAO

Planetario Atlantico
**INTERNATIONAL
PRESS**

Le Monde
La G inuête les services de sécurité européens

Le réquisitoire à huis clos de Loiseau
contre ses homologues européens Renault : l'histoire
de Fiat Chrysler-Nissan-Bercy

Le calvaire de Froome
Infantino : « Ce Mondial
marque un tournant »

LE FIGARO
LE FIGARO LITTÉRAIRE

Frankfurter Allgemeine
ZEITUNG FÜR DEUTSCHLAND

HET LAATSTE NIEUWS

Süddeutsche Zeitung
MÜNCHEN · BERLIN · WÜRZBURG · STUTTGART · LEIPZIG

DIE WELT
SONNENTAG, 12. JUNI 2010

Daily Mail

FINANCIAL TIMES

THE TIMES

How do people browse Wikipedia in response to news events?

Forthcoming PhD submission & articles



News reactions on Wikipedia

- Wiki articles are very different from news articles
- Detect which groups of Wikipedia articles are accessed together in response to a news event
- Persistent groupings across time represent news topics
- Compare to established news agendas





WIKIPEDIA
The Free Encyclopedia

- Main page
- Contents
- Current events
- Random article
- About Wikipedia
- Contact us
- Donate

Contribute

- Help
- Community portal
- Recent changes
- Upload file

Tools

- What links here
- Related changes
- Special pages
- Permanent link
- Page information
- Wikidata item

Print/export

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Portal

Talk

Read

Edit

View history



Portal:Current events/August 2020

From Wikipedia, the free encyclopedia

< [Portal:Current events](#)

2020: [January](#) · [February](#) · [March](#) · [April](#) · [May](#) · [June](#) · [July](#) · **August** · [September](#) · [October](#) · [November](#) · [December](#)

August 2020 is the eighth month of the current leap year. The month, which began on a [Saturday](#), will end on a [Monday](#) after 31 days. It is the current month.

Portal:Current events [edit]

August 1, 2020 (Saturday)

[edit](#) [history](#) [watch](#)

Business and economy

- [China–United States trade war](#)
 - [Chinese](#) IT company [ByteDance](#) agrees to [divest](#) from the U.S. branch of [TikTok](#), in light of [U.S. President Donald Trump's](#) threat to ban the app within the country. If the deal is approved, [Microsoft](#) will handle the app's personal data rather than [ByteDance](#), averting a main source of contention with U.S. officials. [\(Reuters\)](#)^[?]

Disasters and accidents


- [2020 Atlantic hurricane season](#)
 - [Hurricane Isaias](#)
 - [Category One](#) [Isaias](#) makes landfall on northern [Andros Island](#) heading toward [Grand](#)

August 2020						
S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

(transcluded from the [Current events portal](#))

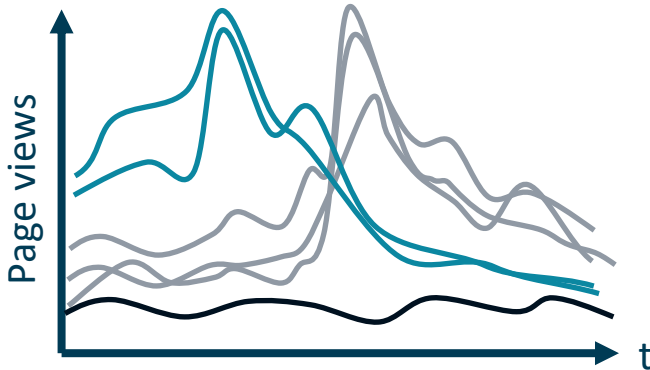
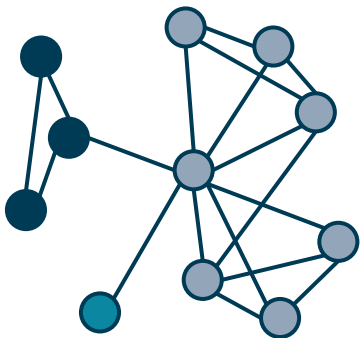
1 year, >4000 events

Extracting article networks

- Can easily get networks of articles for each event using all links
- China–United States trade war
 - Chinese IT company ByteDance agrees to [divest](#) from the U.S. branch of TikTok, in light of U.S. President Donald Trump's threat to ban the app within the country. If the deal is approved, Microsoft will handle the app's personal data rather than ByteDance, averting a main source of contention with U.S. officials. [\(Reuters\)](#) 
 - How to get *meaningful* groups of articles?



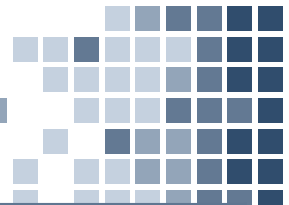
Temporal community detection



- Article network structure
- Long term effects
- Page view correlations
- Short term effects
- Get the groups of articles that are strongly connected AND have similar patterns of page views



Top News Topics on Wikipedia 2018



Topic	Occurrences
US politics	150
Israel-Palestine relations	120
ISIL & Syrian Civil War	64
Trump & family	61
North Korea-South Korea relations	52
War in Afghanistan	41
Yemen Civil War	40
Hurricane & typhoon season	37
Catalan Independence	34
Russian political interference in the US	34

Topic	Excess Page Views
FIFA World Cup	3,241,088
Hurricane & typhoon season	2,808,970
US mass shootings	2,732,908
Skripal poisoning	2,732,084
US politics	2,309,737
Israel-Palestine relations	1,847,287
US govt shutdown	1,749,788
California Wildfires	1,520,025
UK monarchy (Royal Wedding)	1,483,574
US-North Korea relations	1,102,729



Conclusions & Further Work

- Wikipedia is not a news website, but it is an important news information resource
- Patterns of browsing on Wikipedia reflect wider news agenda
- Detecting news topics allows further network-level study e.g. clustering events & topics based on dynamics, and predicting page views
 - (the rest of my PhD!)



Thank you

Any questions?

You can find me at:

@pgildersleve

patrick.gildersleve@oii.ox.ac.uk

References

Dimitrov, D., Singer, P., Lemmerich, F., & Strohmaier, M. (2017, April). What makes a link successful on wikipedia?. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 917-926).

Keegan, B., Gergle, D., & Contractor, N. (2013). Hot off the wiki: Structures and dynamics of Wikipedia's coverage of breaking news events. *American Behavioral Scientist*, 57(5), 595-622.

Miz, V., Benzi, K., Ricaud, B., & Vandergheynst, P. (2017). Wikipedia graph mining: dynamic structure of collective memory. *arXiv preprint arXiv:1710.00398*.

Osborne, M., Petrovic, S., McCreddie, R., Macdonald, C., & Ounis, I. (2012, August). Bieber no more: First story detection using Twitter and Wikipedia. In *Sigir 2012 workshop on time-aware information access*.

Singer, P., Lemmerich, F., West, R., Zia, L., Wulczyn, E., Strohmaier, M., & Leskovec, J. (2017, April). Why we read wikipedia. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 1591-1600).

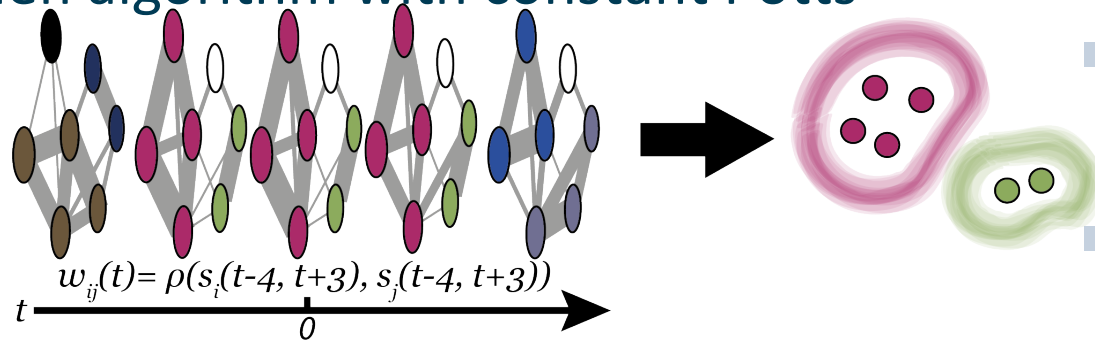
Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1), 1-12.

Presentation template by [SlidesCarnival](#), Photos from Unsplash

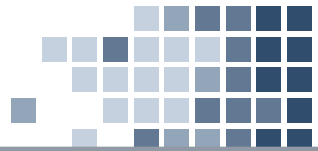


Temporal Community Detection I

- For each news event:
- Use the hyperlink network as a constraint
- Take rolling Pearson correlation of page view time series for linked articles around date of event ► Multilayer network
- Apply Leiden algorithm with constant Potts model



Temporal Community Detection II



- From all events, take all communities
- Construct higher order network
 - Nodes = Communities, Edges = Weighted Jaccard similarity
- Run community detection again on new network
 - Leiden, constant Potts, single layer
- The new clusters represent info topics, apply page view filter to identify news topics

