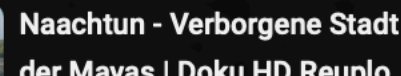


Reducing Exposure to Harmful Content via Graph Rewiring

Corinna Coupette, *Stefan Neumann*, Aristides Gionis

Work currently under submission

February 21, 2023



Djiffmaster



 **Sign in**



Der Kreml ist das herausragende Symbol russischer und sowjetischer Machtentfaltung. Ein architektonisches Ensemble aus Festungsanlagen, Palästen, Kirchen und Regierungsgebäuden. Kein anderes Bauwerk steht mehr für Triumphe und Tragödien Russlands, aber auch für prägende Figuren der Geschichte, große Herrscher, Revolutionäre und Diktatoren.



DIE ERDE ZUR ZEIT DER PANGAEA **1:09:36**



Spreaker

RECHNER-MARKT ON spreaker.com

Kontrafunk

Die Sonntagsrunde mit Burkhard Müller-Ullrich: Blamagenstadt Berlin

KONTRAFUNK

53:51



Der rechte Straßenrand ist Landesgrenze, die in
26:29
200 Metern den Halbeseeherring verlässt.



52:10



3:58:21



Djiffmaster

YouTube

SE

Search

Sign in

Wie ein Mathegenie Hitler knackte | Doku HD | ARTE

Der Untergang der Sowjetunion Doku HD

YouTube

SE

Search

Sign in

YouTube

SE

Search

Sign in

Spreaker

DISCOVER MORE ON [spreaker.com](#)

Kontrafunk

KONTRAFUNK

Die Stimme der Vernunft

Die Sonntagsrunde mit Burkhard Müller-Ullrich: Blamagenstadt Berlin

15K views

11 d

Der Kreml ist da

anderes Bauwer

Wissen-L

1.48K subsc

Superbauten

15,661 views

12 Feb 2023

Die Sonntagsrunde mit Burkhard Müller-Ullrich: Blamagenstadt Berlin

Kontrafunk - Die Stimme der Vernunft

34.5K subscribers

Subscribe

1.5K

Share

Save

Wie ein Mathegenie Hitler knackte | Doku HD | ARTE

ARTEde

501K views · 11 days ago

Musiker und Entertainer Helge Schneider im Interview |...

tagesschau

238K views · 3 days ago

Podcast: ChatGPT und KI - profitieren wirklich alle? | La...

ZDFheute Nachrichten

130K views · 2 days ago

Bericht aus Berlin Extra: Fragen an Sönke Neitzel

tagesschau

118K views · Streamed 7 days ago

Die Maya - Untergang einer Hochkultur | Doku HD...

ARTEde

664K views · 3 weeks ago

Podcast: Waffenlieferungen für die Ukraine - wer sagt wa...

ZDFheute Nachrichten

401K views · 9 days ago

Eine neue Weltordnung: Wie umgehen mit China? | Frank...

ZDFheute Nachrichten

517K views · 7 days ago

Naachtun - Verborgene Stadt der Mayas | Doku HD Reuplo...

ARTEde

581K views · 3 weeks ago

#91 Hat die russische

A collage of overlapping digital interfaces. At the top, a YouTube video player shows a video titled 'Die Sonntagsrunde mit Burkhard Müller-Ullrich: Blamagenstadt Berlin' by 'Kontrafunk - Die Stimme der Vernunft'. The video has 15,661 views and is dated 12 Feb 2023. The description mentions a discussion about the repeat election in Berlin and election fraud. Below the video player is a yellow banner for 'Kontrafunk' with the text 'Die Sonntagsrunde mit Burkhard Müller-Ullrich'. To the right of the banner is a 'Spreaker' logo and a button to 'DISCOVER MORE ON spreaker.com'. Below the banner is a Google Translate interface showing the translation of the video description from German to English. The German text is on the left, and the English translation is on the right. The English translation highlights the phrase 'It's also about the mainstream media's silence on the blatant election fraud scandal'. On the left side of the collage, there are several YouTube thumbnails for various videos, including 'Superbauteile', 'Wissen', and 'Der Kreml ist anders Bau'. On the right side, there are more YouTube thumbnails, including 'Wie ein Mathegenie Hitler knackte | Doku HD | ARTE' and 'Musiker und Entertainer Helge Schneider im Interview |...'. The overall layout is a dense, overlapping composition of digital content.

Send feedback

Die Sonntagsrunde mit Burkhard Müller-Ullrich: Blamagenstadt Berlin

Kontrafunk - Die Stimme der Vernunft
34.5K subscribers

15,661 views 12 Feb 2023
Source:
<https://www.spreaker.com/user/1636937...>

Der Unternehmer und Politiker Marcel Luthe, der Rechtswissenschaftler Ulrich Vosgerau und der Berliner Kontrafunk-Korrespondent Frank Wahlig diskutieren mit Burkhard Müller-Ullrich über die heutige Wiederholungswahl in der deutschen Hauptstadt, die vor dem Bundesverfassungsgericht erzwungen werden musste und trotzdem nur eine halbe Sache ist. Außerdem geht es um das Schweigen der Mainstreammedien zum eklatanten Wahlfälschungsskandal und um die wunderlichen Pläne des ZDF, eine eigene Konkurrenz zu Twitter aufzubauen.

Show less

Spreaker

DISCOVER MORE ON [spreaker.com](https://www.spreaker.com)

Kontrafunk

Die Sonntagsrunde mit Burkhard

Google Translate

Text Documents Websites

GERMAN - DETECTED ENGLISH SPANISH FRENCH

ENTREPRENEUR AND POLITICIAN MARCEL LUTHE, LEGAL SCHOLAR ULRICH VOSGERAU AND BERLIN KONTRAFUNK CORRESPONDENT FRANK WAHLIG DISCUSS TODAY'S REPEAT ELECTION IN THE GERMAN CAPITAL WITH BURKHARD MÜLLER-ULLRICH, WHICH HAD TO BE FORCED BEFORE THE FEDERAL CONSTITUTIONAL COURT AND IS STILL ONLY A HALF THING. IT'S ALSO ABOUT THE MAINSTREAM MEDIA'S SILENCE ON THE BLATANT ELECTION FRAUD SCANDAL AND ZDF'S WHIMSICAL PLANS TO BUILD UP ITS OWN COMPETITION TO TWITTER.

520 / 5,000

Following the
recommendations
has led us to
alt-right content!

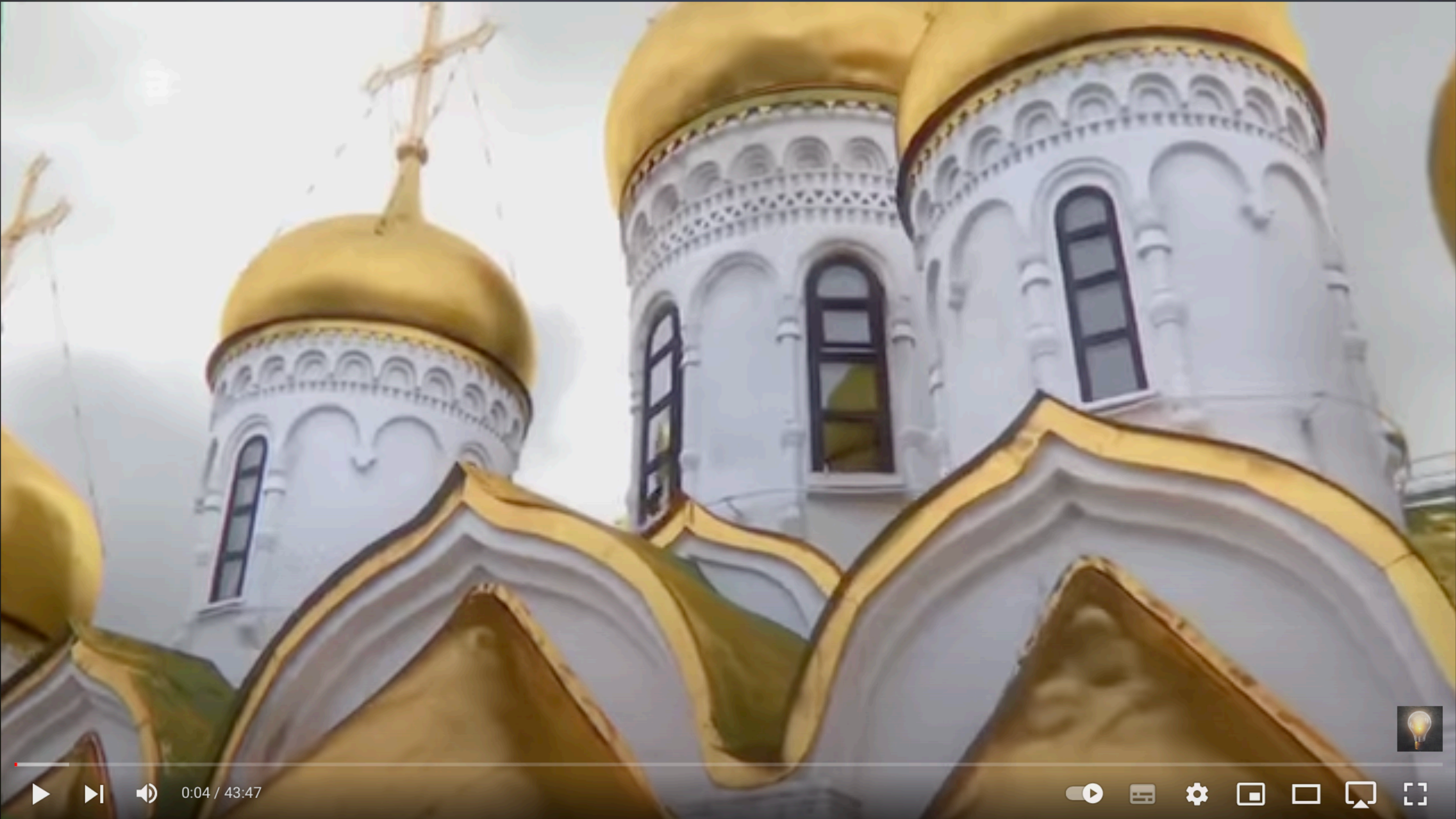
YouTube

SE

Search

Q

Sign in




Superbauten der Geschichte - Der Kreml Doku HD


Wissen-Doku-Info
1.48K subscribers

15K views 11 days ago


Der Kreml ist das herausragende Symbol russischer und sowjetischer Machtentfaltung. Ein architektonisches Ensemble aus Festungsanlagen, Palästen, Kirchen und Regierungsgebäuden. Kein anderes Bauwerk steht mehr für Triumphe und Tragödien Russlands, aber auch für prägende Figuren der Geschichte, große Herrscher, Revolutionäre und Diktatoren.




Der Untergang der Sowjetunion Doku HD
Wissen-Doku-Info
4.9K views · 1 month ago




Wie sah die Erde zur Zeit von Pangaea aus? | Dokumentar...
Modysee | Die Welt der Odysseen
3.7K views · 6 hours ago




Die Sonntagsrunde mit Burkhard Müller-Ullrich:...
Kontrafunk - Die Stimme der Vernunft
15K views · 18 hours ago




Honeckers Regierungsbunker Doku HD
Wissen-Doku-Info
22K views · 5 months ago




Die Sieben Geheimnisse der NVA DOKU HD
Wissen-Doku-Info
115K views · 3 months ago




Kleinlichtenhain - ein Niemandsland (?) an der...
SRF- Südthüringer Regionalfernsehen
44K views · 2 weeks ago



Stalins deutsche Elite | Deutsche Spezialisten im...
Doku Kanal - Militär und Geschichte
27K views · 4 months ago




10 000 Jahre vor Geheimnisse der alten Zivilisationen (Doku...
Hörspiel - Sci-Fi - Krimi - Horror
37K views · 8 days ago



DOKU - Es geschah in NRW Der große Benzinbetrug -...
Djiffmaster

We can represent the recommendations as a **graph**

 **Der Untergang der Sowjetunion Doku HD**
Wissen-Doku-Info
59:32 4.9K views · 1 month ago

 **Wie sah die Erde zur Zeit von Pangaea aus? | Dokumentar...**
Modysee | Die Welt der Odysseen
1:09:36 3.7K views · 6 hours ago
New

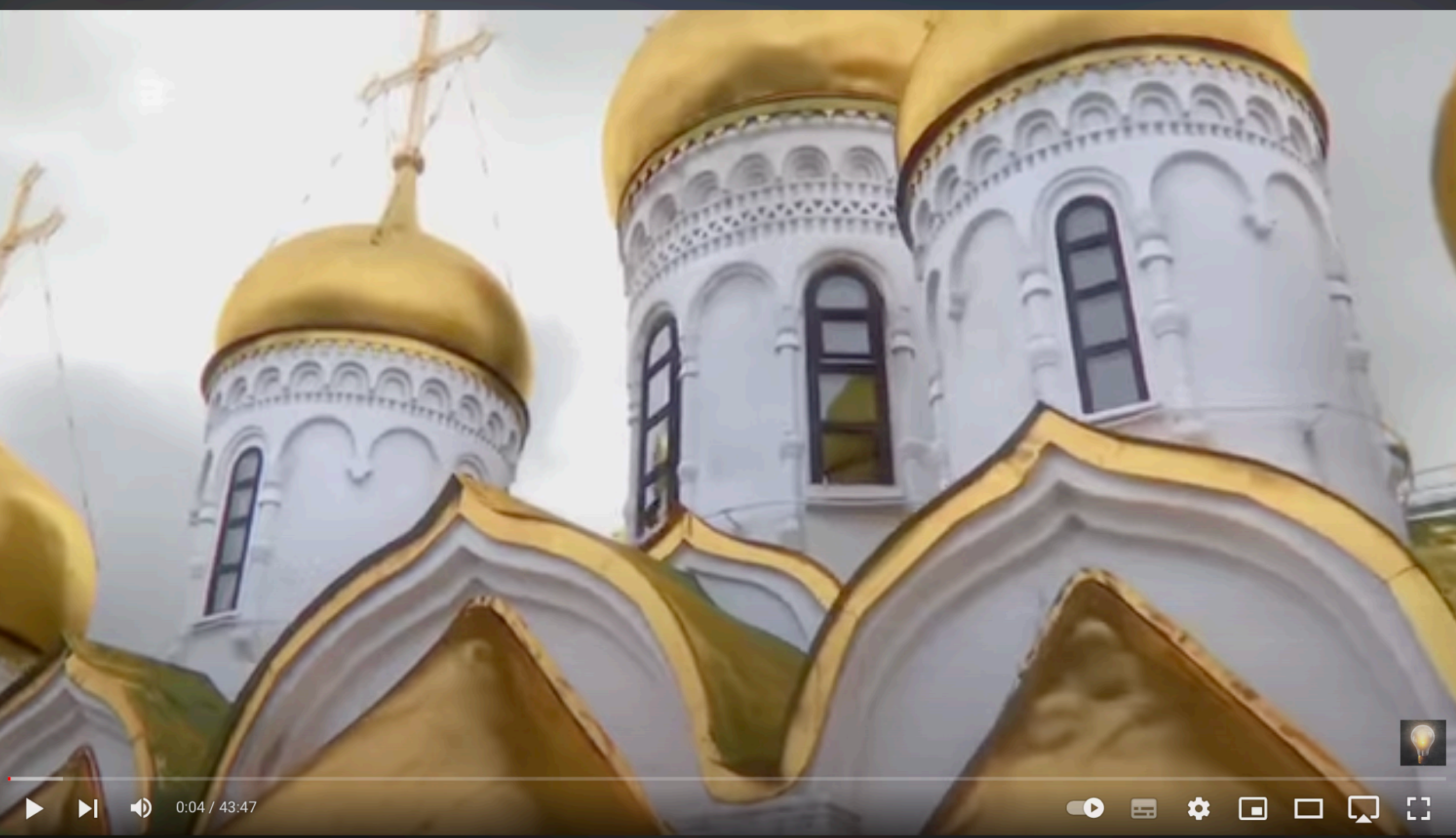
 **Die Sonntagsrunde mit Burkhard Müller-Ullrich:...**
Kontrafunk - Die Stimme der Vernunft
53:51 15K views · 18 hours ago
New

 **Honeckers Regierungsbunker Doku HD**
Wissen-Doku-Info
44:04 22K views · 5 months ago


 **Die Sieben Geheimnisse der NVA DOKU HD**
Wissen-Doku-Info
44:49 115K views · 3 months ago

 **Kleinlichtenhain - ein Niemandsland (?) an der...**
SRF- Südthüringer Regionalfernsehen
26:29 44K views · 2 weeks ago


 **Stalins deutsche Elite | Deutsche Spezialisten im...**
Doku Kanal - Militär und Geschichte
52:10 27K views · 4 months ago



Superbauten der Geschichte - Der Kreml Doku HD

 **Wissen-Doku-Info**
1.48K subscribers

[Subscribe](#)

144  [Share](#) [Save](#) ...

15K views 11 days ago

Der Kreml ist das herausragende Symbol russischer und sowjetischer Machtentfaltung. Ein architektonisches Ensemble aus Festungsanlagen, Palästen, Kirchen und Regierungsgebäuden. Kein anderes Bauwerk steht mehr für Triumphe und Tragödien Russlands, aber auch für prägende Figuren der Geschichte, große Herrscher, Revolutionäre und Diktatoren.





Superbauten der Geschichte - Der Kreml Doku HD

Wissen-Doku-Info
1.48K subscribers

Subscribe

144

Share

Save

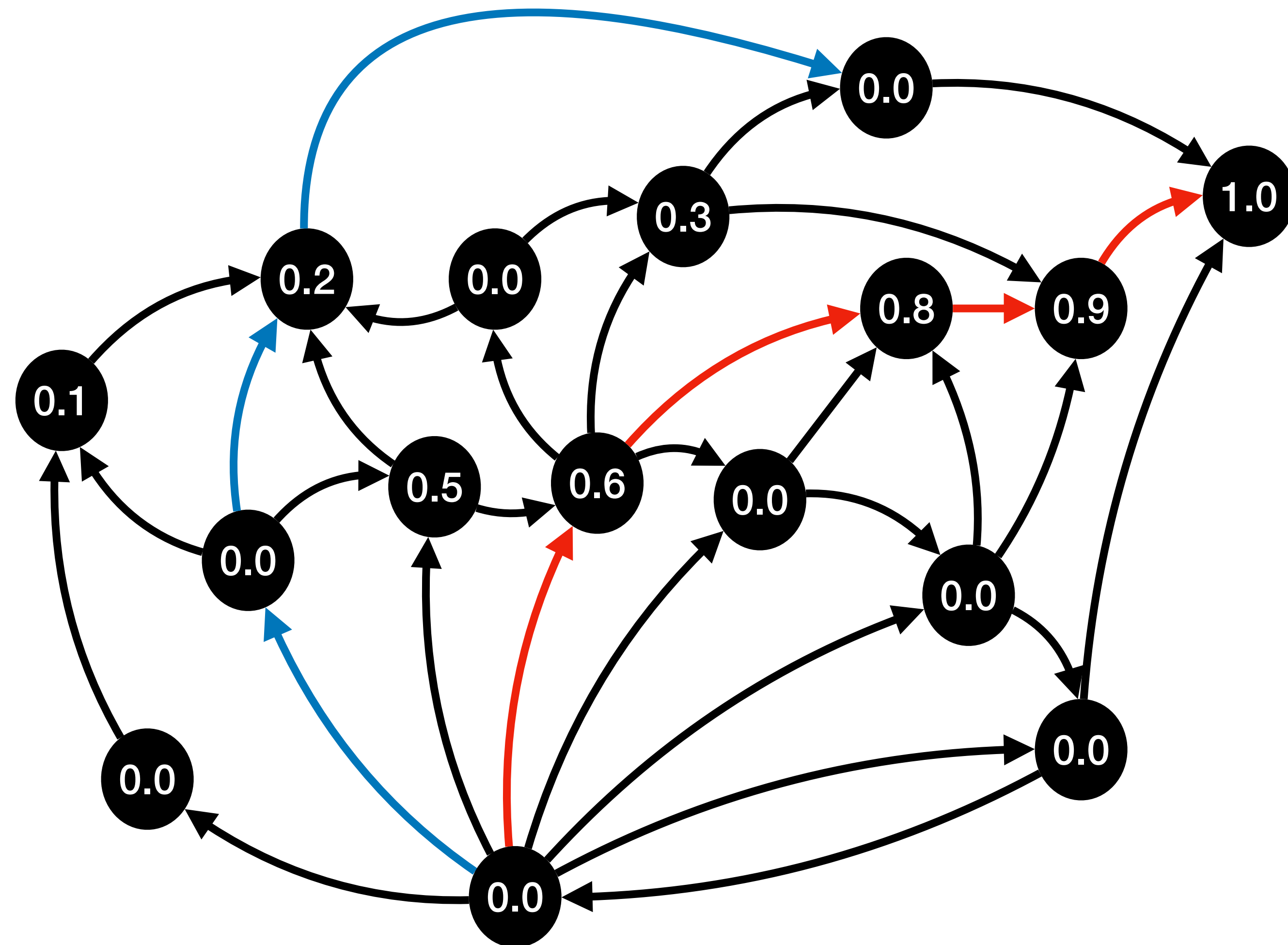
15K views 11 days ago

Der Kreml ist das herausragende Symbol russischer und sowjetischer Machtentfaltung. Ein architektonisches Ensemble aus Festungsanlagen, Palästen, Kirchen und Regierungsgebäuden. Kein anderes Bauwerk steht mehr für Triumphe und Tragödien Russlands, aber auch für prägende Figuren der Geschichte, große Herrscher, Revolutionäre und Diktatoren.



Recommendation Graph

- We can model the recommendations as a graph:
 - Nodes ~ videos
 - Edge (i, j) ~ video j recommended when watching video i
 - Each node i has a harmfulness score c_i in $[0, 1]$
 - ➡ 0 = not harmful,
1 = very harmful

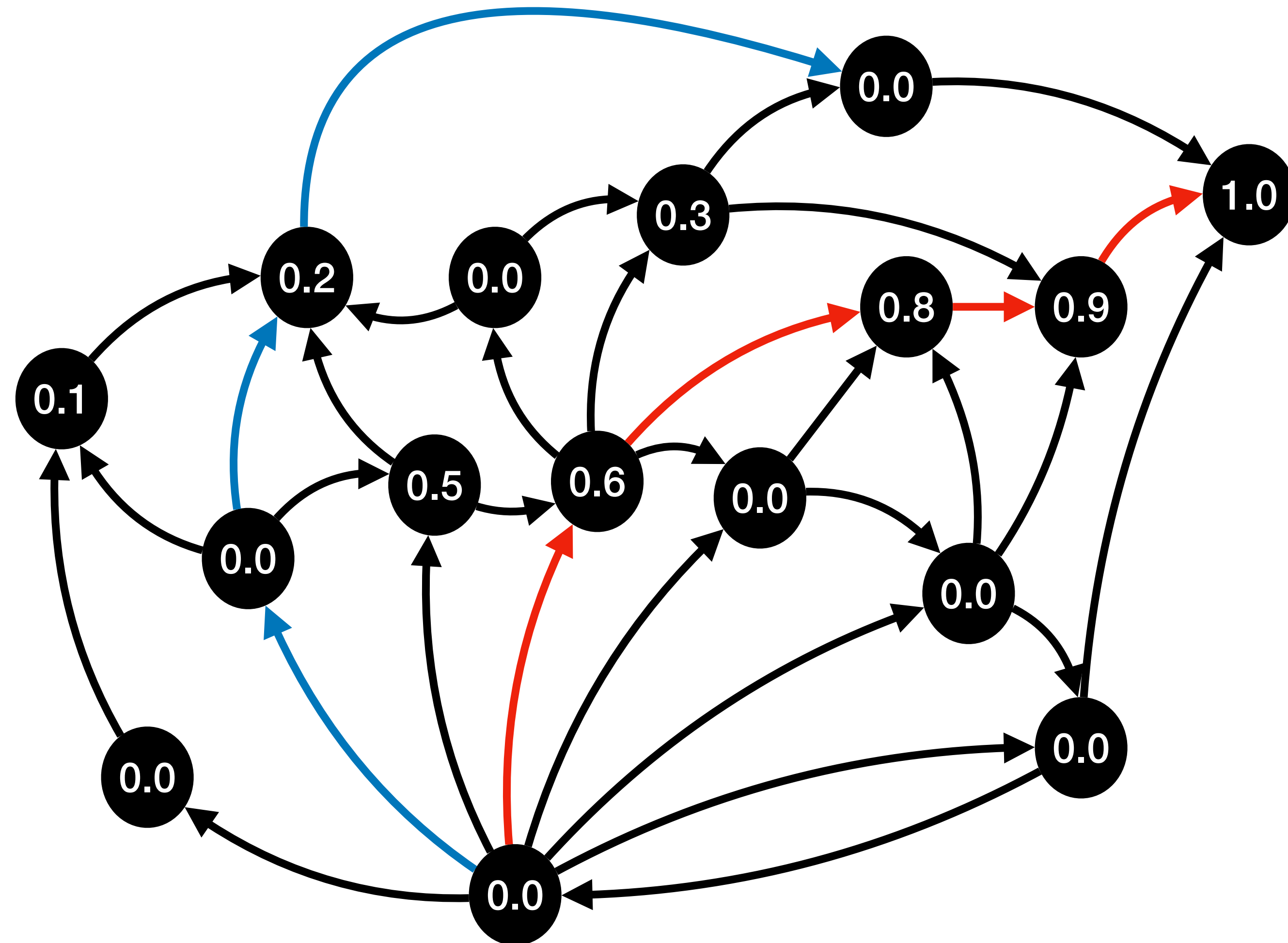


blue path: harmfulness 0.2

red path: harmfulness 3.3

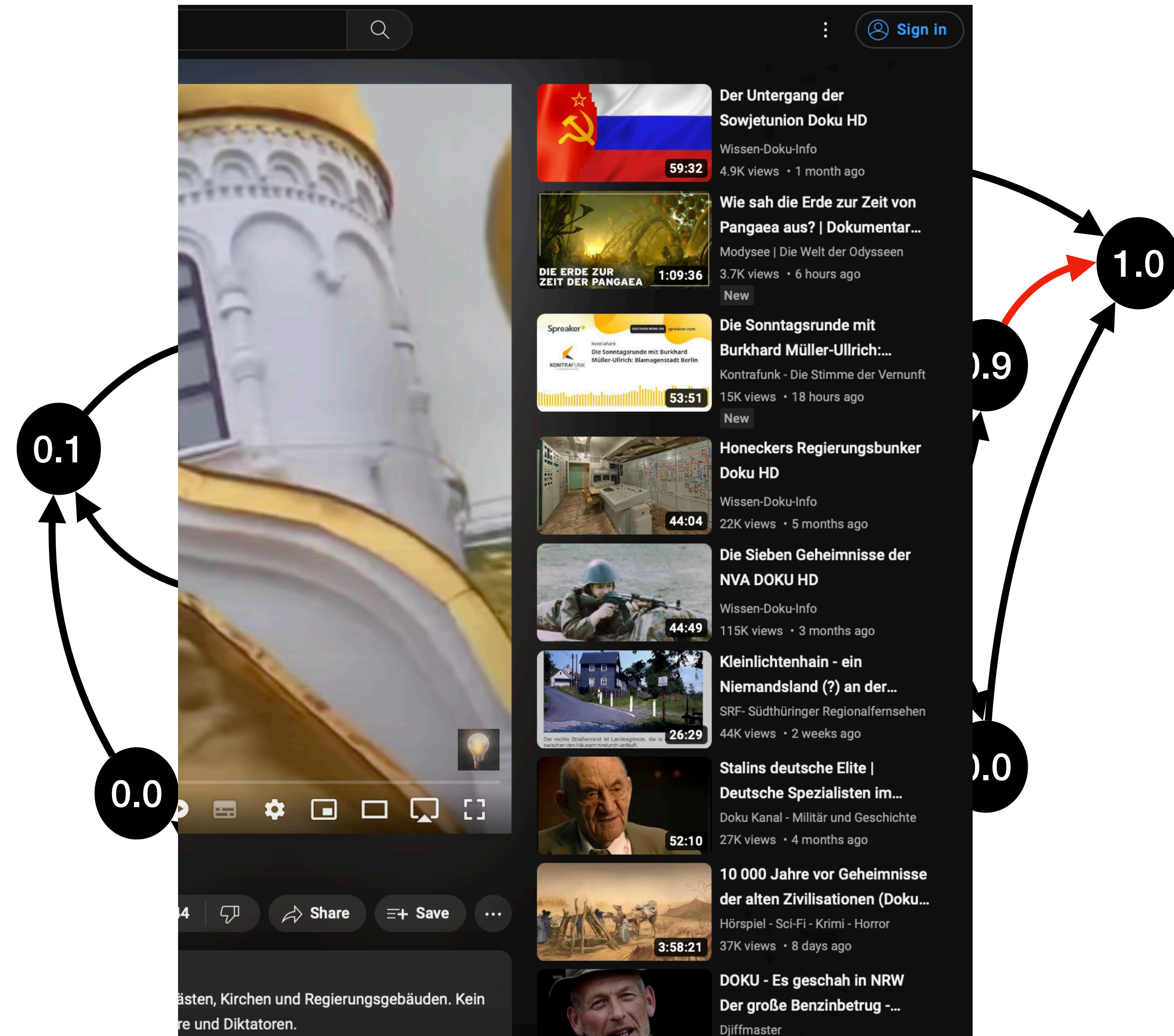
Problem Statement

- **Measuring harmfulness:**
Start random walks at all nodes and compute expected exposure to harm
- **Problem statement:**
Perform r rewirings such that the total exposure to harmfulness is minimized
 - “Rewiring”: remove edge (i, j) and add edge (i, k)
 - Corresponds to removing the recommendation of a harmful video and replacing it with a less harmful one



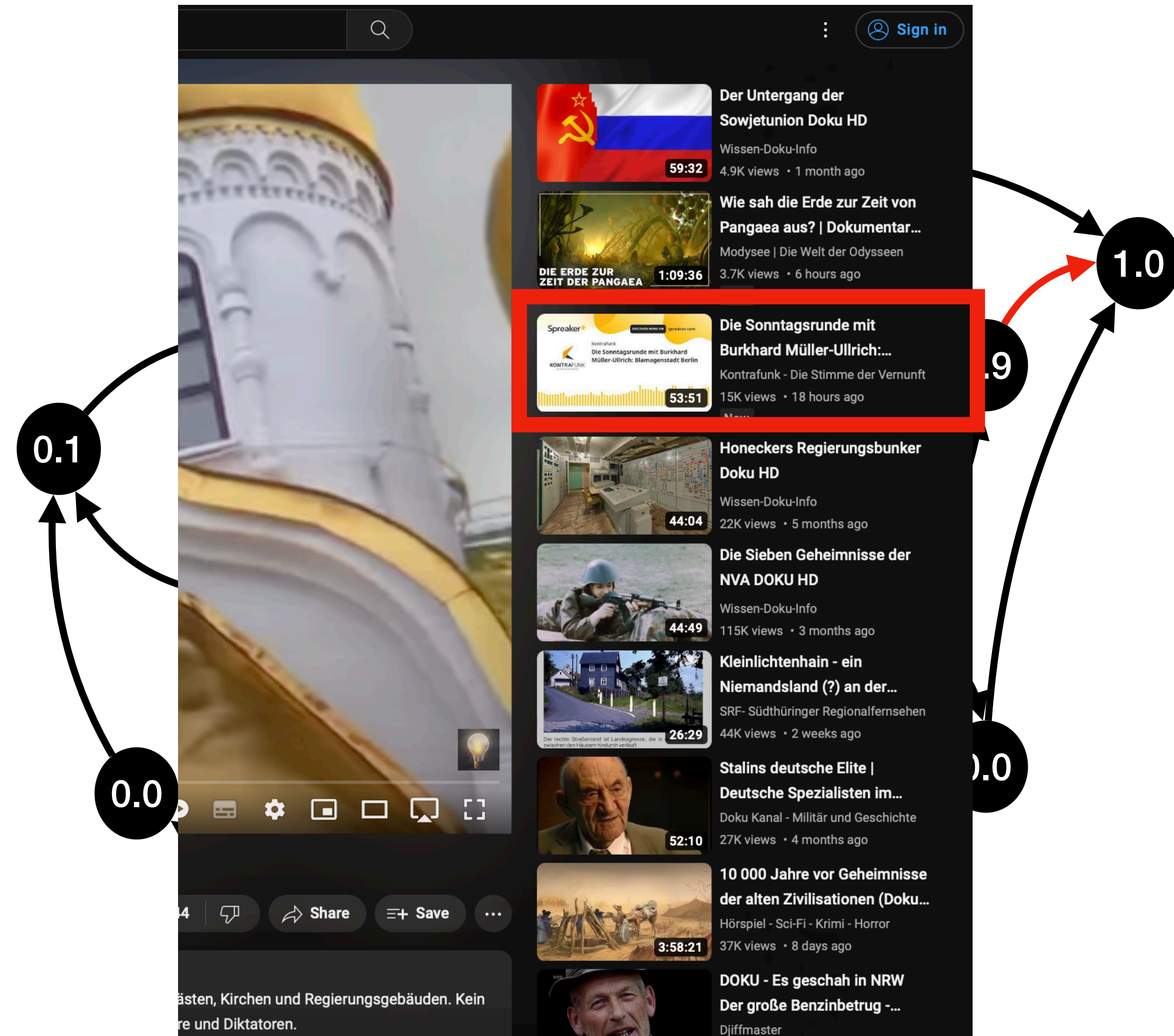
Problem Statement

- **Measuring harmfulness:**
Start random walks at all nodes and compute expected exposure to harm
- **Problem statement:**
Perform r rewirings such that the total exposure to harmfulness is minimized
 - “Rewiring”: remove edge (i, j) and add edge (i, k)
- Corresponds to removing the recommendation of a harmful video and replacing it with a less harmful one



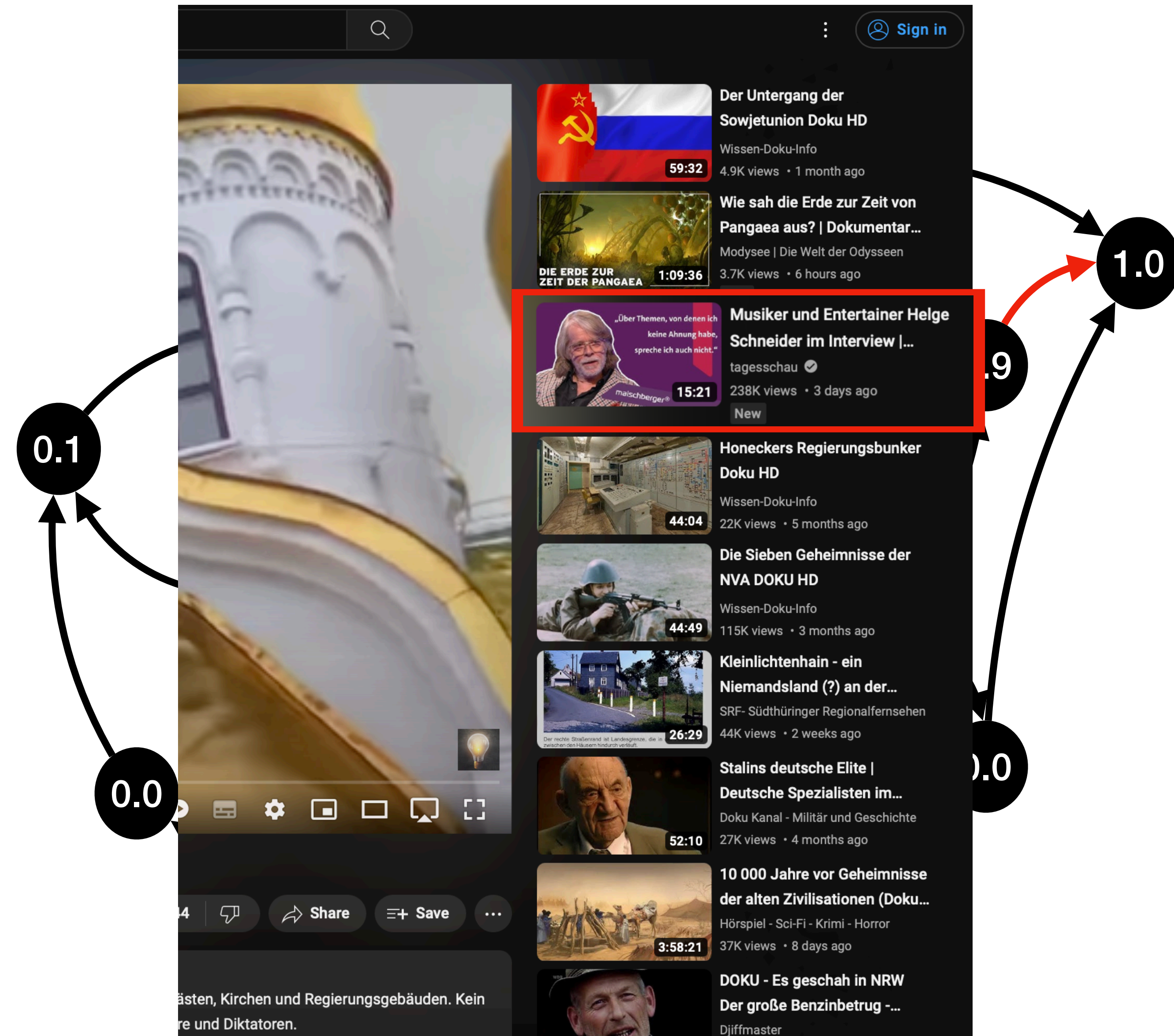
Problem Statement

- **Measuring harmfulness:**
Start random walks at all nodes and compute expected exposure to harm
- **Problem statement:**
Perform r rewirings such that the total exposure to harmfulness is minimized
 - “Rewiring”: remove edge (i, j) and add edge (i, k)
 - Corresponds to removing the recommendation of a harmful video and replacing it with a less harmful one



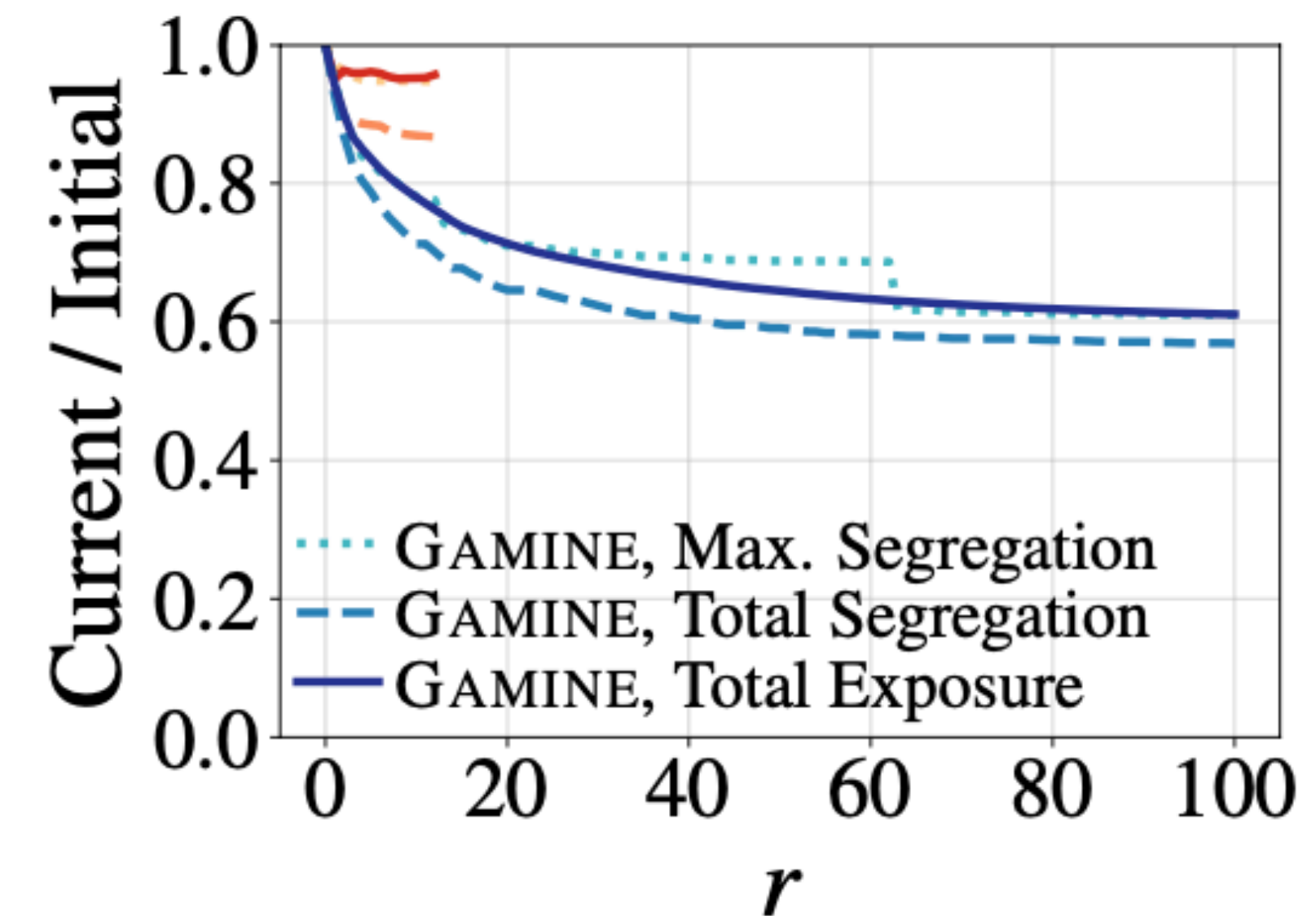
Problem Statement

- **Measuring harmfulness:**
Start random walks at all nodes and compute expected exposure to harm
- **Problem statement:**
Perform r rewirings such that the total exposure to harmfulness is minimized
 - “Rewiring”: remove edge (i, j) and add edge (i, k)
 - Corresponds to removing the recommendation of a harmful video and replacing it with a less harmful one



Results

- We provide an algorithm that can significantly reduce exposure to harmful content
 - ➡ Under mild assumptions, provides 0.63-approximation
 - ➡ We prove NP-hardness
- On real-world YouTube datasets, we can significantly reduce harmful content even with ≤ 100 rewirings



(b) YT-10K, $q = 0.99$

Data

- Synthetic data (random graphs)
- YouTube dataset:
 - Real-world YouTube recommendation graphs, harmfulness scores based on channel categories
 - “Available upon request” by Ribeiro et al. (FAT*, 2020), shared via Dropbox
 - After receiving the data, we stored and anonymized it in the same way as if we had created the dataset ourselves
- NELA-GT-2021:
 - News articles with harmfulness score based on Media Bias/Fact Check
 - Publicly available, created by Gruppi et al. (arxiv, 2021)

Ethical Considerations

- **Intended usage:**
 - Our algorithm's goal is to reduce exposure to harmful contents
 - Changing only “a few” recommendation is a milder intervention than censoring harmful content directly
- **Potential intended abuse:**
 - When using manipulated cost function for harmful contents, our algorithm can be used to discriminate against contents, e.g., by the political opposition
 - However, the same manipulation could be made to the existing recommendation systems that are being deployed in practice
- **Potential unintended side effects:**
 - Deploying our algorithm in practice might lead to unexpected side effects
 - Can be prevented by rigorous impact assessments and cost function audits before and during deployment

A ETHICS STATEMENT

In this work, we introduce GAMINE, a method to reduce the exposure to harm induced by recommendation algorithms on digital media platforms via edge rewiring, i.e., replacing certain recommendations by others. While removing harm-inducing recommendations constitutes a milder intervention than censoring content directly, it still steers attention away from certain content to other content, which, if pushed to the extreme, can have censorship-like effects. Although in its intended usage, GAMINE primarily counteracts the tendency of recommendation algorithms to overexpose harmful content as similar to other harmful content, when fed with a contrived cost function, it could also be used to discriminate against content considered undesirable for problematic reasons (e.g., due to political biases or stereotypes against minorities). However, as the changes to recommendations suggested by GAMINE could also be made by amending recommendation algorithms directly, the risk of *intentional* abuse is no greater than that inherent in the recommendation algorithms themselves, and *unintentional* abuse can be prevented by rigorous impact assessments and cost function audits before and during deployment. Thus, we are confident that overall, GAMINE can contribute to the health of digital platforms.

Appendix

Exposure to Harmful Content

- Random walks correspond to users who follow the recommendations

- Consider walk $w = (i = u_0, u_1, \dots, u_k)$

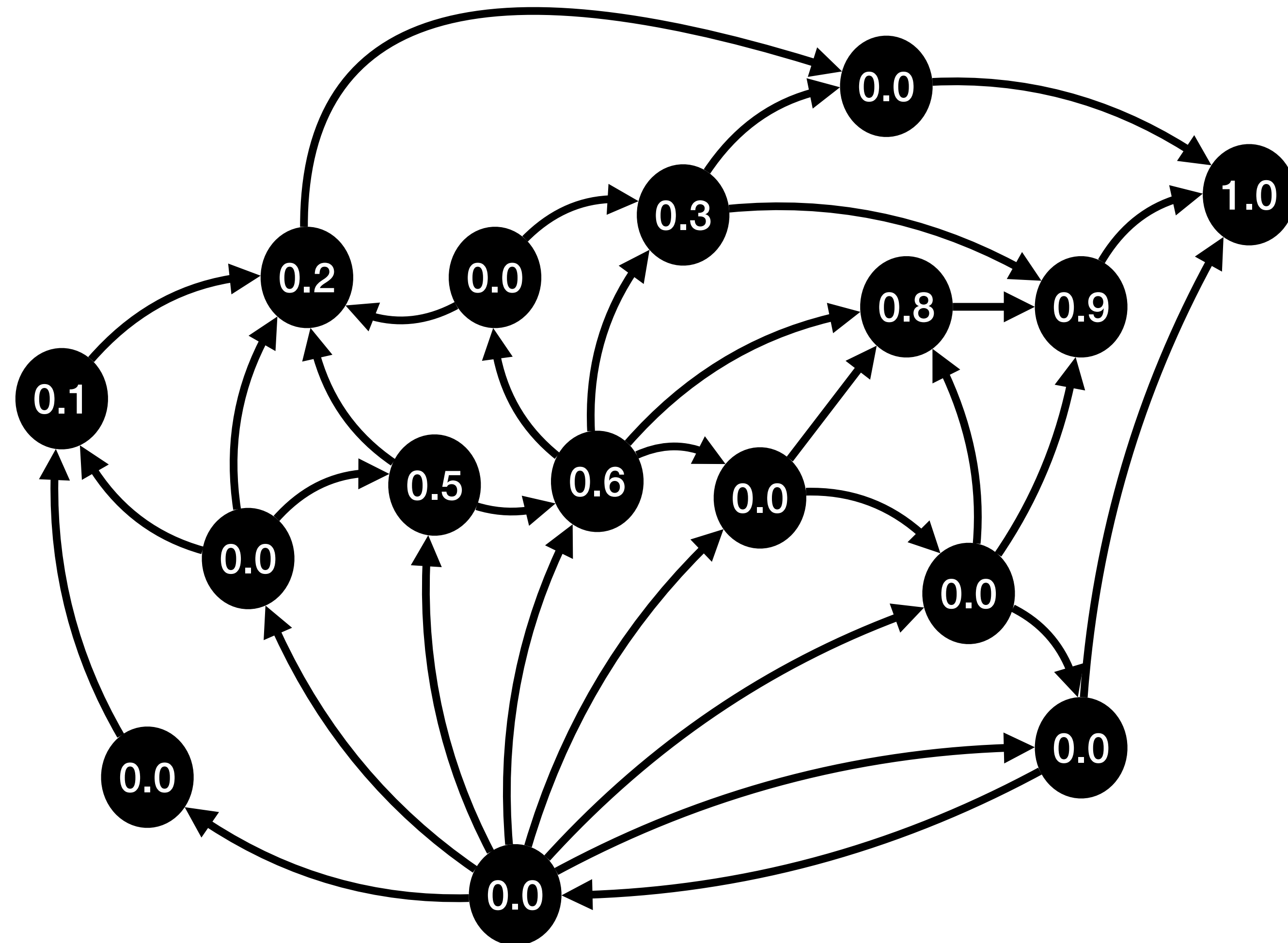
- Harmfulness of walk w starting at node i :

$$H_{i,w} = \sum_i c_{u_i}$$

- For each video, with probability α we stop the random walk

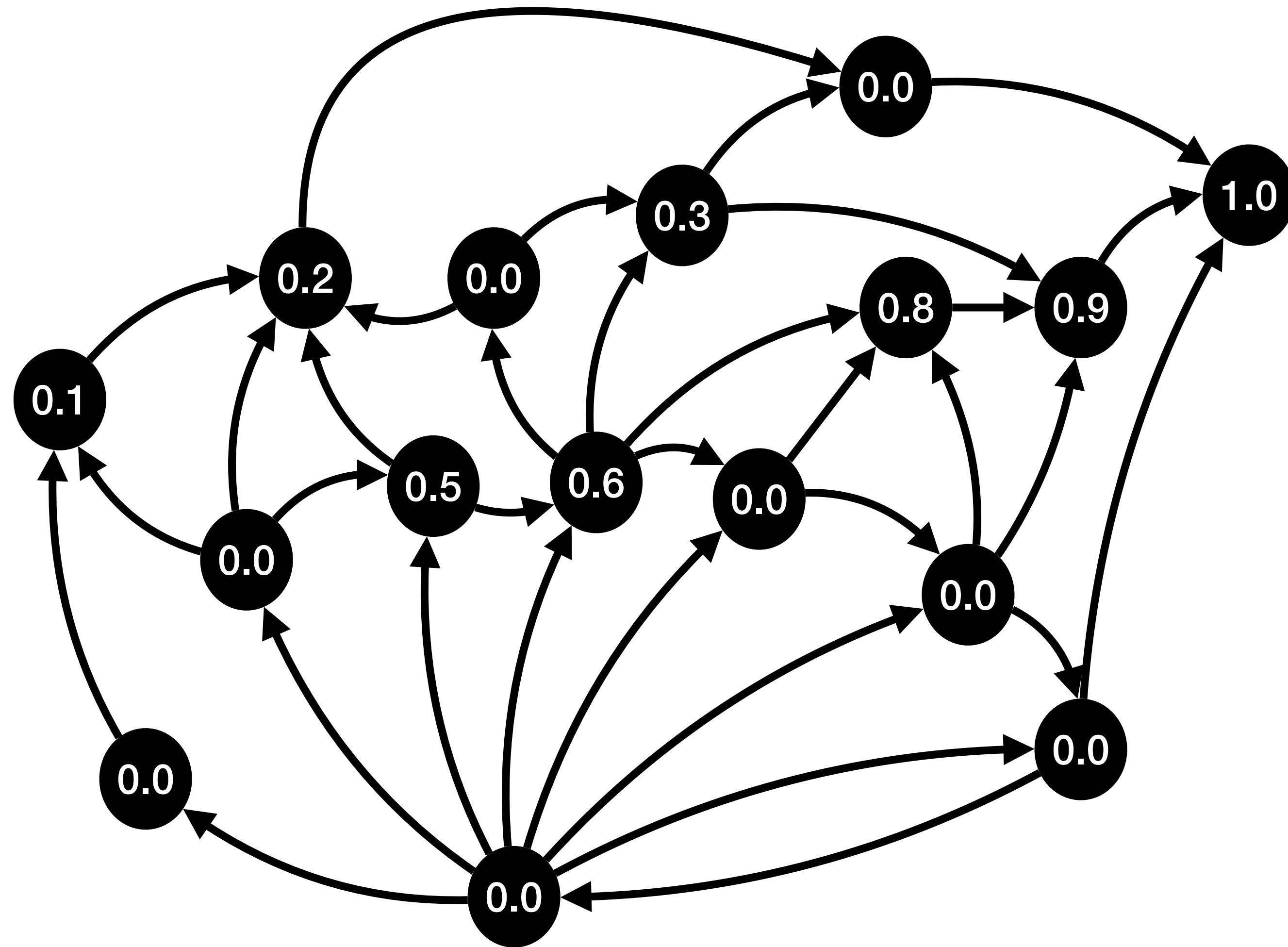
- Exposure of user i is given by $\mathbb{E}_w[H_{i,w}]$, where the randomness is taken over all random walks w

- Total exposure for network: $\sum_i \mathbb{E}_w[H_{i,w}]$



Problem Statement

- Perform k rewirings such that the total exposure for the network $\sum_i \mathbb{E}_w[H_{i,w}]$ is minimized
- “Rewiring”: remove edge (i, j) and add edge (i, k)
- Corresponds to removing the recommendation of a harmful video and replacing it with a less harmful one



Harmfulness Scores for YouTube Dataset

- anti-feminist channels (Incel, MGTOW, MRA, and PUA)

Table 7: Costs of videos from each channel under our four different cost functions.

Category	c_{B1}	c_{B2}	c_{R1}	c_{R2}
Alt-lite	1.0	1.0	0.8	0.8
Alt-right	1.0	1.0	1.0	1.0
Incel	0.0	1.0	0.4	0.6
IDW	1.0	1.0	0.6	0.2
MGTOW	0.0	1.0	0.4	0.6
MRA	0.0	1.0	0.2	0.4
NONE	0.0	0.0	0.0	0.0
PUA	0.0	1.0	0.2	0.4
center	0.0	0.0	0.0	0.0
left	0.0	0.0	0.0	0.0
left-center	0.0	0.0	0.0	0.0
right	0.0	0.0	0.0	0.0
right-center	0.0	0.0	0.0	0.0