Coursework and financial network examples

LEARNING OUTCOMES

To learn about how I want you to do the coursework so you don’t get disappointed

Real-world applications of network science in finance

ACADEMIC MISCONDUCT

Please remember the good scholarly practice requirements of the University regarding work for credit. You can find guidance at the School page

https://web.inf.ed.ac.uk/infweb/admin/policies/academic-misconduct

This also has links to the relevant University pages.

ACADEMIC MISCONDUCT

Please remember the good scholarly practice requirements of the University regarding work for credit. You can find guidance at the School page

READ THIS VERY WELL!!!

https://web.inf.e[illegible]mic-misconduct

This also has links to the relevant University pages.

SUBMISSION

-All the analysis **must** be done in Python.
-You **must** submit all your code in a single zip file (even if it is just a single python file, it must be zipped).
-You also **must** submit a pdf in which you report the results and discuss them.

Example:
S123456789.zip
S123456789.pdf

HOW IT WORKS

Your boss/manager/customers will **not tell you what to do**, step by step

You need to learn how to become **independent** in the analysis of networks

You will be able to stay up to date and provide useful insights only if you **know what you are doing**

COURSEWORK

Main goal: Study a competitor through network analysis

Dataset:

Projected bipartite network of funds' portfolios

Available on https://opencourse.inf.ed.ac.uk/dbba/assessment

COURSEWORK

You have been **hired as a data analyst** in the newly founded investment company DBBA Capital and have been tasked with the **analysis of the investment patterns of one of our major competitors**: Fairholme Capital, managed by Bruce Berkowitz.

COURSEWORK

DBBA Capital wants you **to evaluate the investment patterns of Fairholme Capital** in relation to other superinvestors and evaluate the change in investment patterns during the pandemic. They have provided you with data about different superinvestors and the companies they invested in for each quarter spanning from quarter 1 (Q1) of 2019 to quarter 2 (Q2) of 2023 (that you can find in the folder named "Assignment Data").

COURSEWORK

TIP: When you believe it might help, **make use of the information you have on the portfolio composition** to comment and discuss your results.

COURSEWORK

TASK 1.1 (7 marks)

Load the first Excel dataset ("2019_Q1.xlsx") and create a network out of the investors and companies in the following manner:

- the nodes of the network are all the investors in the first column of the dataset
- two investors (nodes) are connected with an edge if the have invested in the same company (e.g. Christopher Bloomstran - Semper Augustus and David Abrams - Abrams Capital Management will be connected because they both invested in GOOGL).
- if two investors have invested in more than one common company, **do *not* assign multiple edges** between them. Instead, assign the number of common companies they have invested in as a weight to the edge connecting them.

After you built the network, extract the **largest connected component** and plot it. Remember to add the edge weights in your plot.

COURSEWORK

TASK 1.2 (3 marks)

Obtain the ego-network of 'Bruce Berkowitz - Fairholme Capital' and plot it.

COURSEWORK

Task 2.1 (15 marks)

Now that you know how to build the network for a single quarter and get its largest connected component, repeat the procedure for all the other quarters. For both the whole network and the ego-network, **produce a table with the summary statistics** (i.e. mean, max, min, and standard deviation) of the following network quantities:

- Number of nodes
- Number of links
- Density
- Average clustering coefficient
- Average degrees
- Average strength
- Assortativity

If you need to make any **assumption** or decision regarding the metric to use to compute any of these quantities, clearly **motivate** it.

COURSEWORK

	Mean	Median	Min.	Max.	SD		Mean	Median	Min	Max	SD
Payments						Distance Measures					
Volume (,000)	436	411	371	644	60.3	$\langle \ell \rangle$	2.62	2.63	2.56	2.66	0.02
Valuc ($tr)	1.30	1.27	1.13	1.64	0.11	$\langle \varepsilon \rangle$	4.67	4.63	4.18	5.74	0.33
Average ($mn)	3.01	3.06	2.48	3.35	0.20	D	6.6	7	6	7	0.5
Components						$M(2)$ (%)	41.6	41.3	38.9	47.3	2.0
GWCC	6,460	6,484	6,355	6,729	83	$M(3)$ (%)	95.9	95.8	95.1	97.1	0.5
DC	2	2	0	8	2	$M(4)$ (%)	99.9	99.9	99.8	100	0.0
GSCC (n)	5,086	5,066	4,914	5,395	123	Clustering					
GIN	527	528	404	645	49	$\langle C \rangle$	0.53	0.53	0.51	0.55	0.01
GOUT	774	782	595	916	67	Degree Distribution					
Tendrils	103	103	88	116	7	$\langle k \rangle$	15.2	14.8	13.9	17.6	0.8
Connectivity and Reciprocity						Max k^{out}	1,922	1,913	1,772	2,269	121
m	76,614	75,397	69,077	94,819	6,151	Max k^{in}	2,097	2,070	1,939	2,394	115
p (%)	0.3	0.29	0.28	0.33	0.01	$\hat{\gamma}^{out}_{\mathrm{MLE}}$	2.11	2.11	2.09	2.14	0.01
r (%)	21.5	21.5	21	23	0.03	$\hat{\gamma}^{in}_{\mathrm{MLE}}$	2.15	2.15	2.15	2.18	0.01

TABLE II: Turnover, component and network statistics for the Fedwire interbank payment network, fourth quarter 2004. $tr = $trillion, $mn = $million, GWCC = giant weakly connected component, GSCC = giant strongly connected component, GIN = giant in component, GOUT = giant out component, DC = Disconnected component. All network statistics are calculated for GSCC. n = size, m = number of links, p = connectivity, r = reciprocity, $\langle \ell \rangle$ = average path length, $\langle \varepsilon \rangle$ = average eccentricity, D = diameter, $M(x)$ = mass distance function, $\langle C \rangle$ = clustering coefficient, $\langle k \rangle$ = average degree, k^{in} = in-degree, k^{out} = out-degree, γ = power law coefficient.

COURSEWORK

Task 2.2 (10 marks)

Discuss why ego networks are useful for exploring the importance of singular nodes. Then, comment on the statistics you computed above and what information they give you about the investment patterns of Bruce Berkowitz - Fairholme Capital. Briefly discuss how the ego network statistics differ from the statistics obtained for the whole network, explaining whether the differences or similarities are expected or not. **Motivate your answers.**

COURSEWORK

Task 2.2 (10 marks)

Discuss why ego networks are useful for exploring the importance of singular nodes. Then, comment on the statistics you computed above and what information they give you about the investment patterns of Bruce Berkowitz - Fairholme Capital. Briefly discuss how the ego network statistics differ from the statistics obtained for the whole network, explaining whether the differences or similarities are expected or not. **Motivate your answers.**

COURSEWORK

Task 3.1 (8 marks)

Choose a single temporal slice (i.e. quarter) and plot and analyse the total degree and strength distributions of both the whole network and the ego-network. Comment on the similarities/differences between these networks.

COURSEWORK

Task 3.1 (8 marks)

Choose a single temporal slice (i.e. quarter) and plot and **analyse** the total degree and strength distributions of both the whole network and the ego-network. **Comment** on the similarities/differences between these networks.

COURSEWORK

Task 3.2 (7 marks)

Based on degree distributions **and the results you obtained**, what type of network would you say the whole network and ego-network are (e.g scale free, random, etc)? Could have they been generated by any of the models discussed in class? **Motivate your answer.**

COURSEWORK

Task 4.1 (15 marks)

Plot the **temporal evolution** of the quantities you computed in Part 2 for the ego network and the whole network compare the difference between the networks. For each quantity, **discuss** if it can be used for analysing the investment patterns of Bruce Berkowitz - Fairholme Capital over time. Based on your discussion, **choose the quantities that you find important**. What information you can draw about the change of those network statistics during the pandemic?

COURSEWORK

Tasks 4.2 (10 marks)

Choose a **suitable centrality measure** that would give us important information about the nodes in the whole network, and **clearly motivate your choice**. Use this measure to find the 3 most central nodes for each quarter. Compare the centrality of Bruce Berkowitz - Fairholme Capital overtime with that of the most central nodes. What can you conclude from this?

COURSEWORK

Task 5.1 (15 marks)

Find the communities in each quarter in the whole network. To do so, use an algorithm of your choice, and **justify your decision**. Analyse how the communities evolve overtime, **focussing on the membership of Bruce Berkowitz - Fairholme Capital**. Does this node fall in the same community with the same superinvestors across different quarters? What **conclusions** can you draw from this?

COURSEWORK

Task 6.1 (10 marks)

As any good DBBA Capital data analyst, at the end of your analysis you need to present your findings. Please write a brief (~250 words) report discussing how the portfolio of Fairholme Capital has changed compared with the rest of the funds in the dataset.

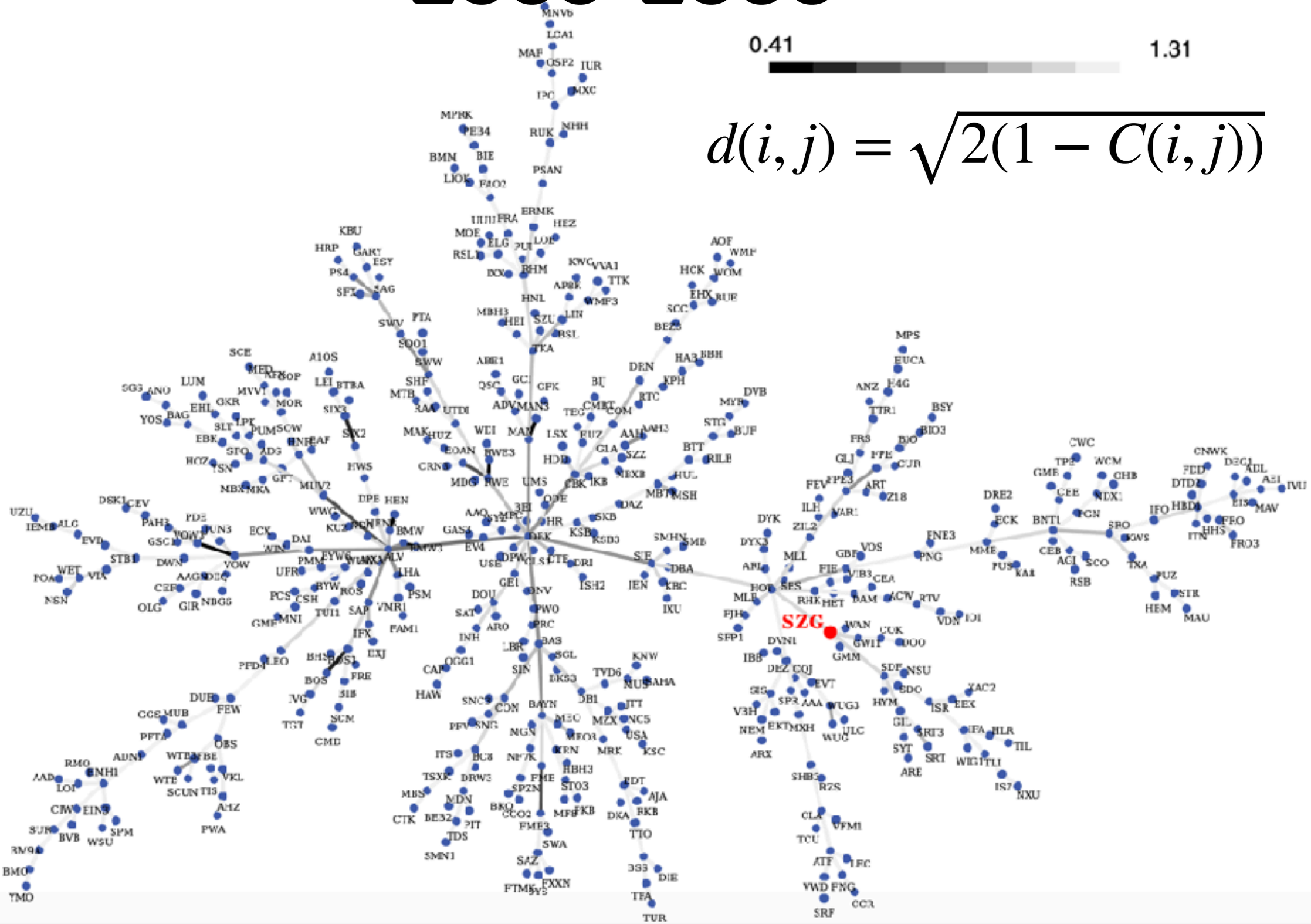
INTERNSHIP OPPORTUNITIES

CASE STUDY I – HOW TO BECOME MILLIONAIRES WITH NETWORKS

THIS PAPER STUDIES THE TOPOLOGICAL FEATURES OF THE **CORRELATION NETWORK** OF THE FRANKFURT STOCK EXCHANGE (FSE)
THE AUTHORS SHOW THAT THERE ARE **PHASE TRANSITIONS**
BEFORE AND AFTER THE 2008 **CRISIS**

M. WILINSKI ET AL. **"STRUCTURAL AND TOPOLOGICAL PHASE TRANSITIONS ON THE GERMAN EXCHANGE STOCK MARKET"** https://arxiv.org/pdf/1301.2530.pdf

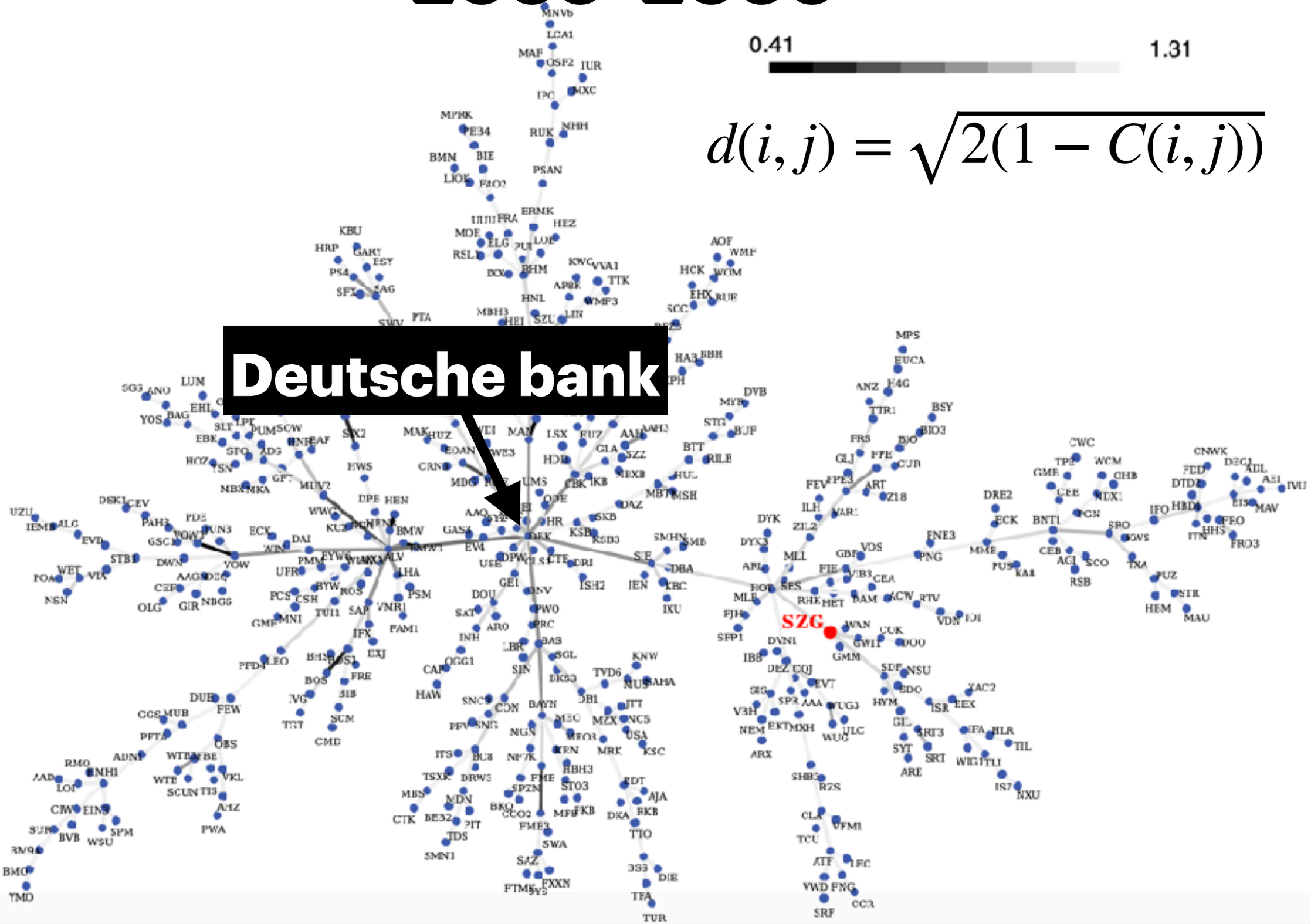
2005-2006



$$d(i,j) = \sqrt{2(1 - C(i,j))}$$

M. WILINSKI ET AL. **“STRUCTURAL AND TOPOLOGICAL PHASE TRANSITIONS ON THE GERMAN EXCHANGE STOCK MARKET”** https://arxiv.org/pdf/1301.2530.pdf

2005-2006

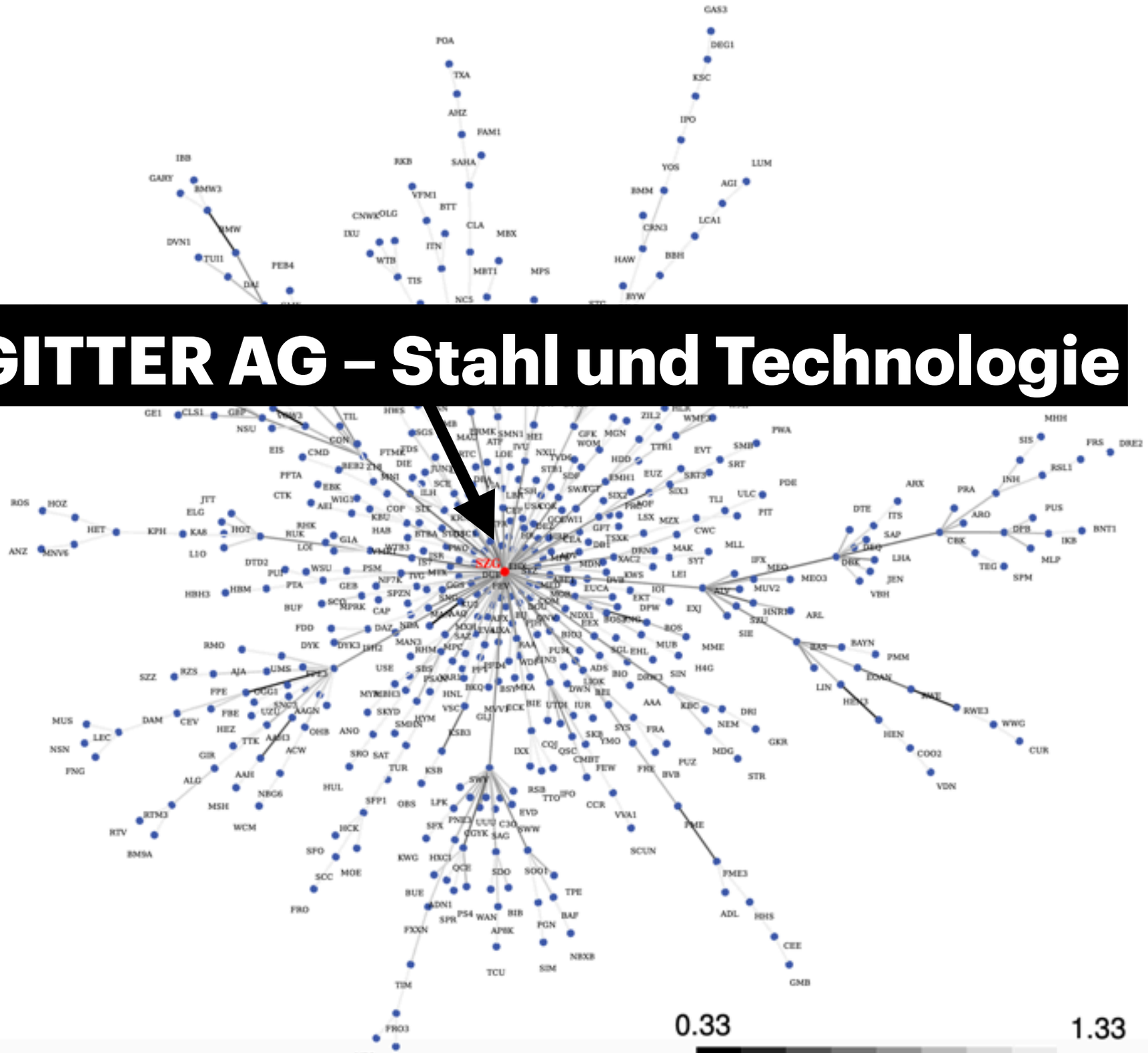


M. WILINSKI ET AL. **“STRUCTURAL AND TOPOLOGICAL PHASE TRANSITIONS ON THE GERMAN EXCHANGE STOCK MARKET”** https://arxiv.org/pdf/1301.2530.pdf

2005-2006

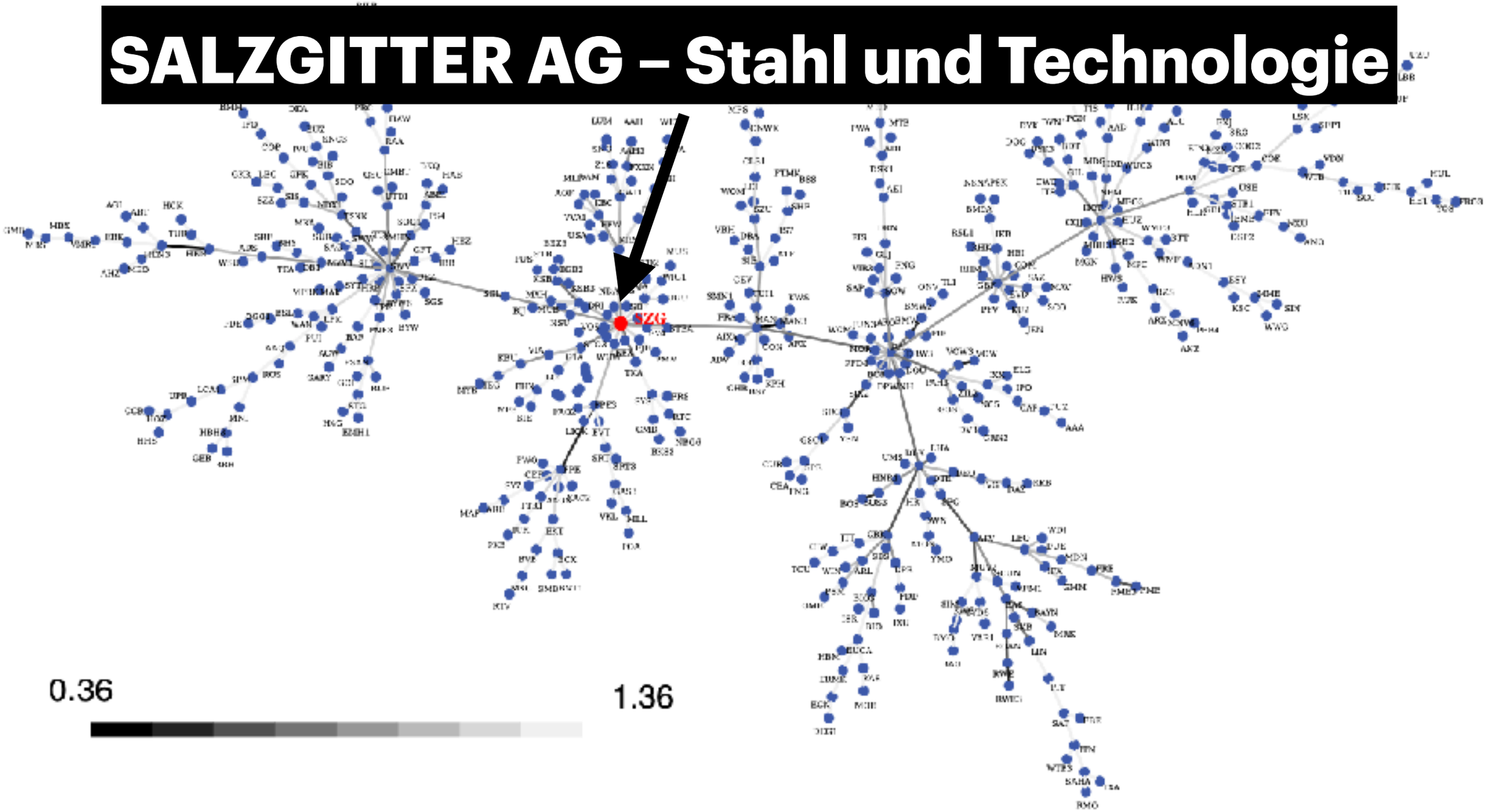
M. WILINSKI ET AL. **“STRUCTURAL AND TOPOLOGICAL PHASE TRANSITIONS ON THE GERMAN EXCHANGE STOCK MARKET”** https://arxiv.org/pdf/1301.2530.pdf

2006-2007



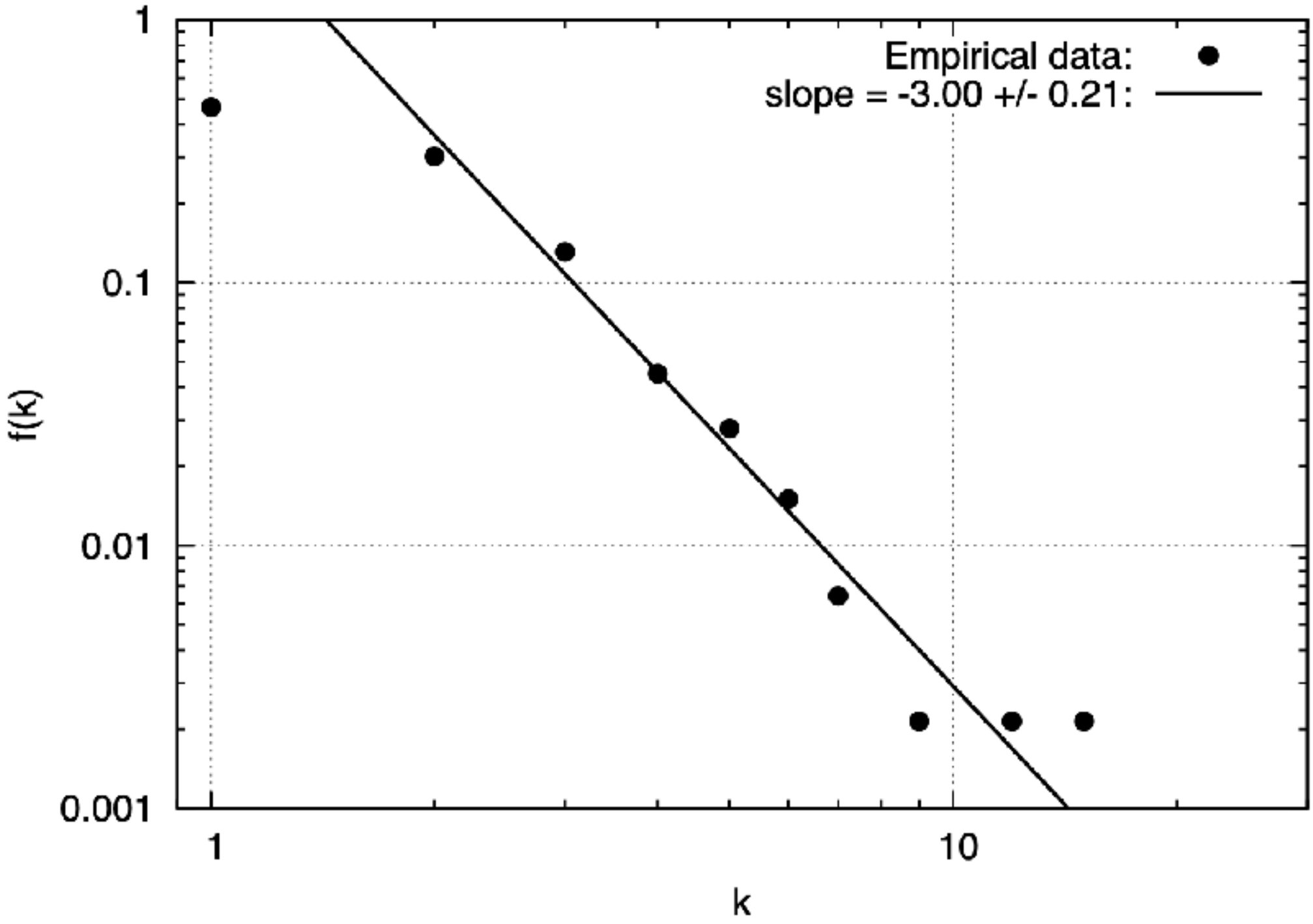
M. WILINSKI ET AL. **“STRUCTURAL AND TOPOLOGICAL PHASE-TRANSITIONS ON THE GERMAN EXCHANGE STOCK MARKET”** https://arxiv.org/pdf/1301.2530.pdf

2007-2008



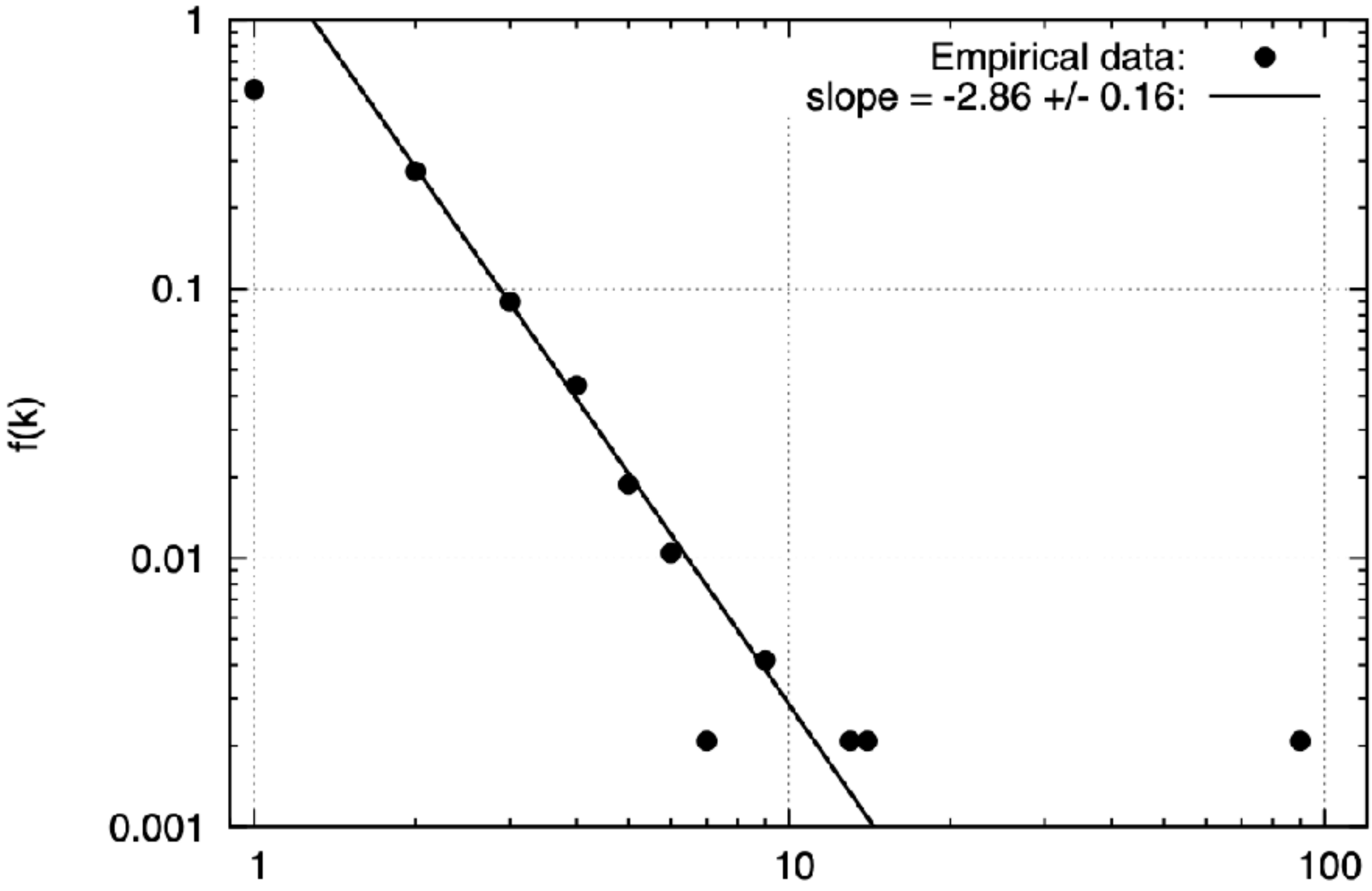
M. WILINSKI ET AL. **“STRUCTURAL AND TOPOLOGICAL PHASE TRANSITIONS ON THE GERMAN EXCHANGE STOCK MARKET”** https://arxiv.org/pdf/1301.2530.pdf

2005-2006



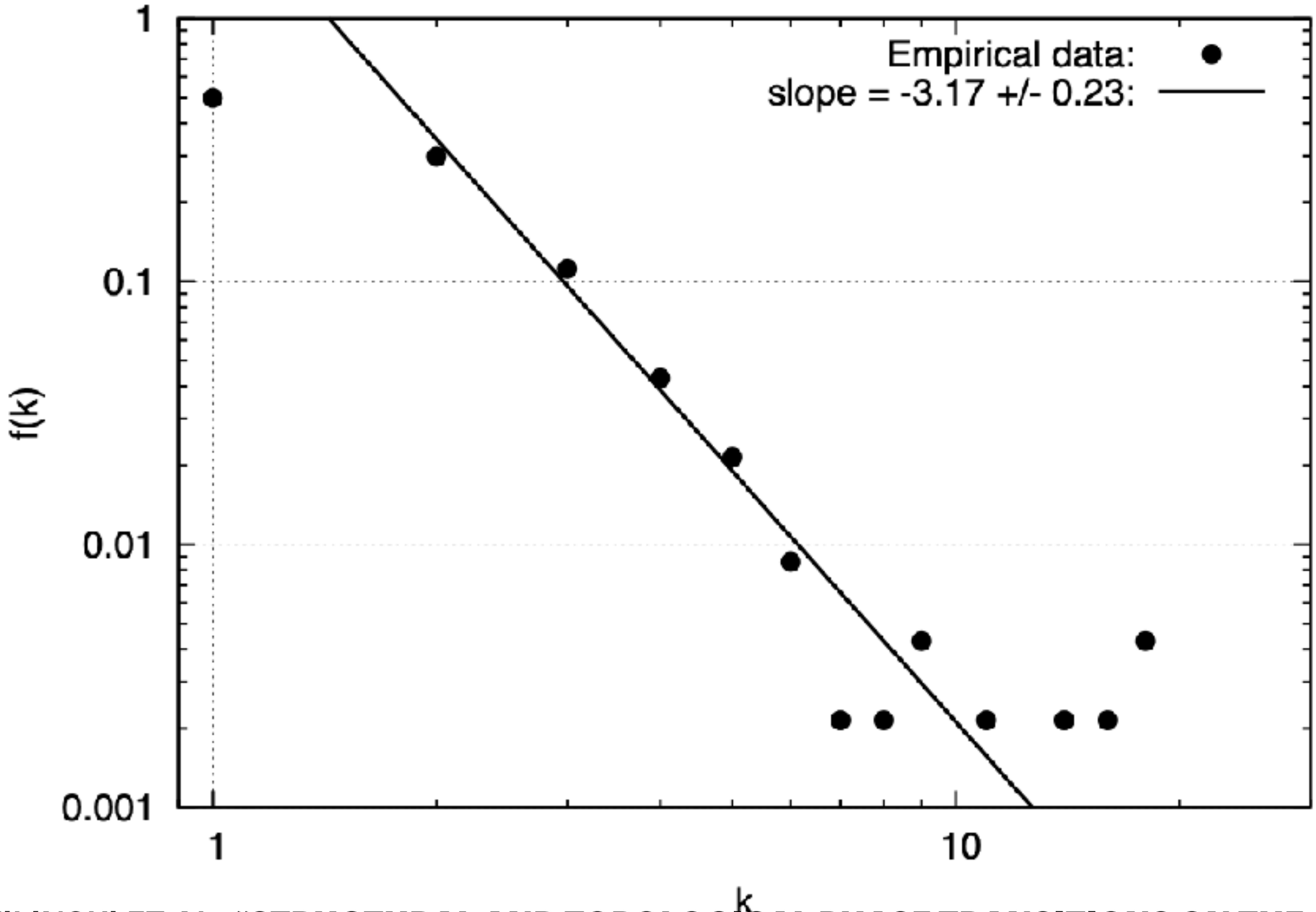
M. WILINSKI ET AL. **“STRUCTURAL AND TOPOLOGICAL PHASE TRANSITIONS ON THE GERMAN EXCHANGE STOCK MARKET”** https://arxiv.org/pdf/1301.2530.pdf

2006-2007



M. WILINSKI ET AL. **“STRUCTURAL AND TOPOLOGICAL PHASE TRANSITIONS ON THE GERMAN EXCHANGE STOCK MARKET”** https://arxiv.org/pdf/1301.2530.pdf

2007-2008



M. WILINSKI ET AL. **“STRUCTURAL AND TOPOLOGICAL PHASE TRANSITIONS ON THE GERMAN EXCHANGE STOCK MARKET”** https://arxiv.org/pdf/1301.2530.pdf

05/06: Phase of scale-free MST - a (relatively) stable stock market state

06/07: Phase of the superstar-like MST - a transient market state

07/08: Phase of scale-free MST decorated by few local star-like trees - a (relatively) stable stock market state

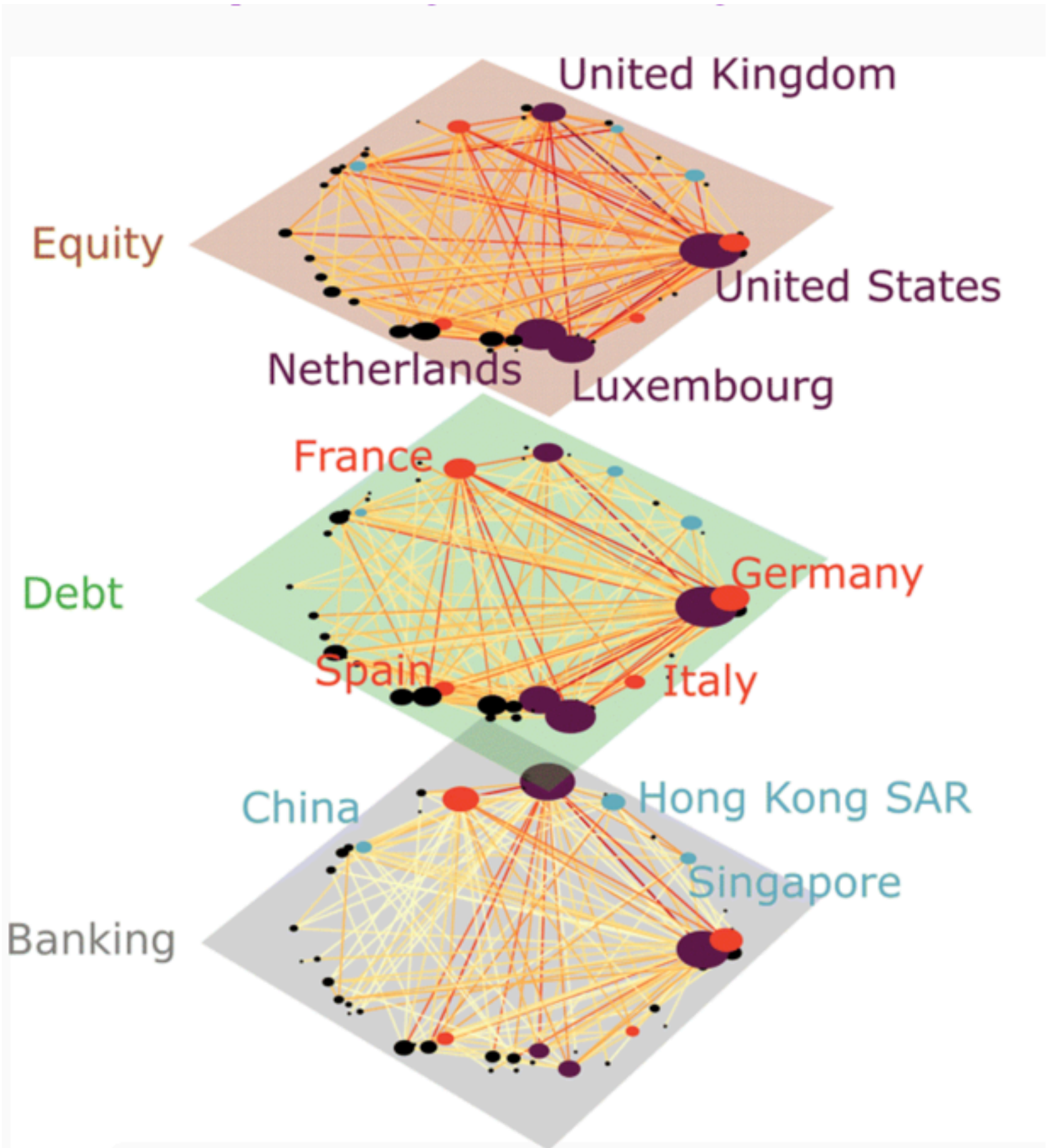
M. WILINSKI ET AL. **"STRUCTURAL AND TOPOLOGICAL PHASE TRANSITIONS ON THE GERMAN EXCHANGE STOCK MARKET"** https://arxiv.org/pdf/1301.2530.pdf

CASE STUDY II – FINANCIAL CONTAGION

The authors analyse financial contagion using **multilayer networks**

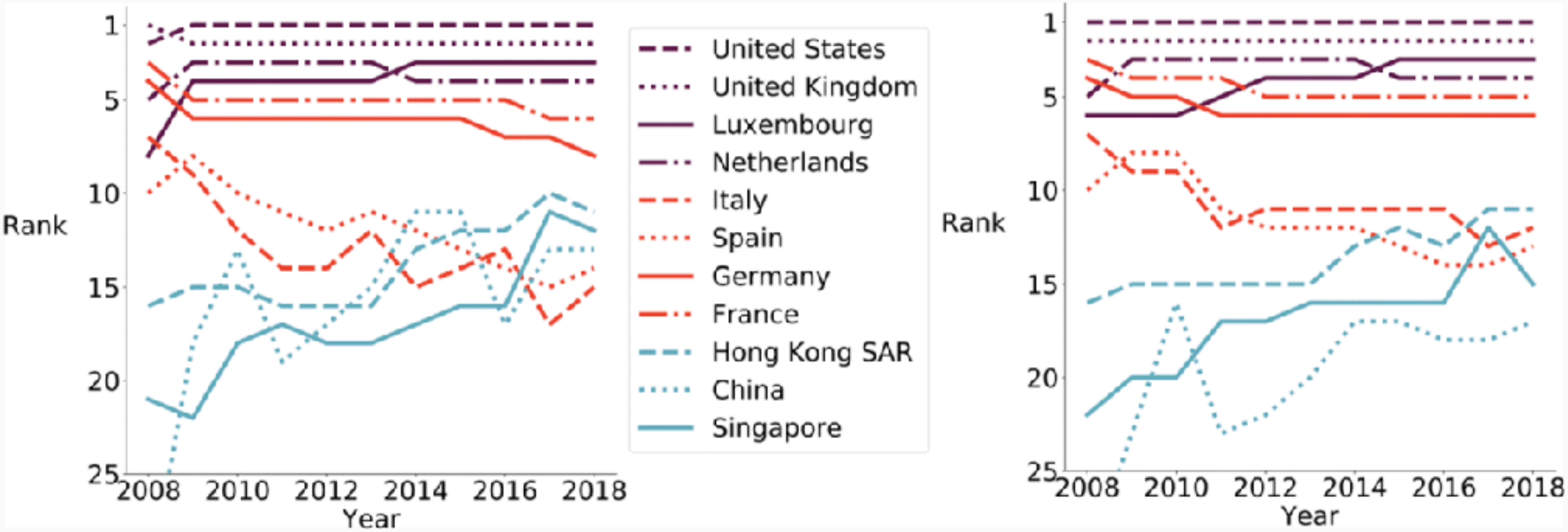
They find that using multilayer networks, they can find up to **2x important countries for contagion**

M. DEL RIO-CHANONA ET AL. **"THE MULTIPLEX NATURE OF GLOBAL FINANCIAL CONTAGIONS"**
https://appliednetsci.springeropen.com/articles/10.1007/s41109-020-00301-2

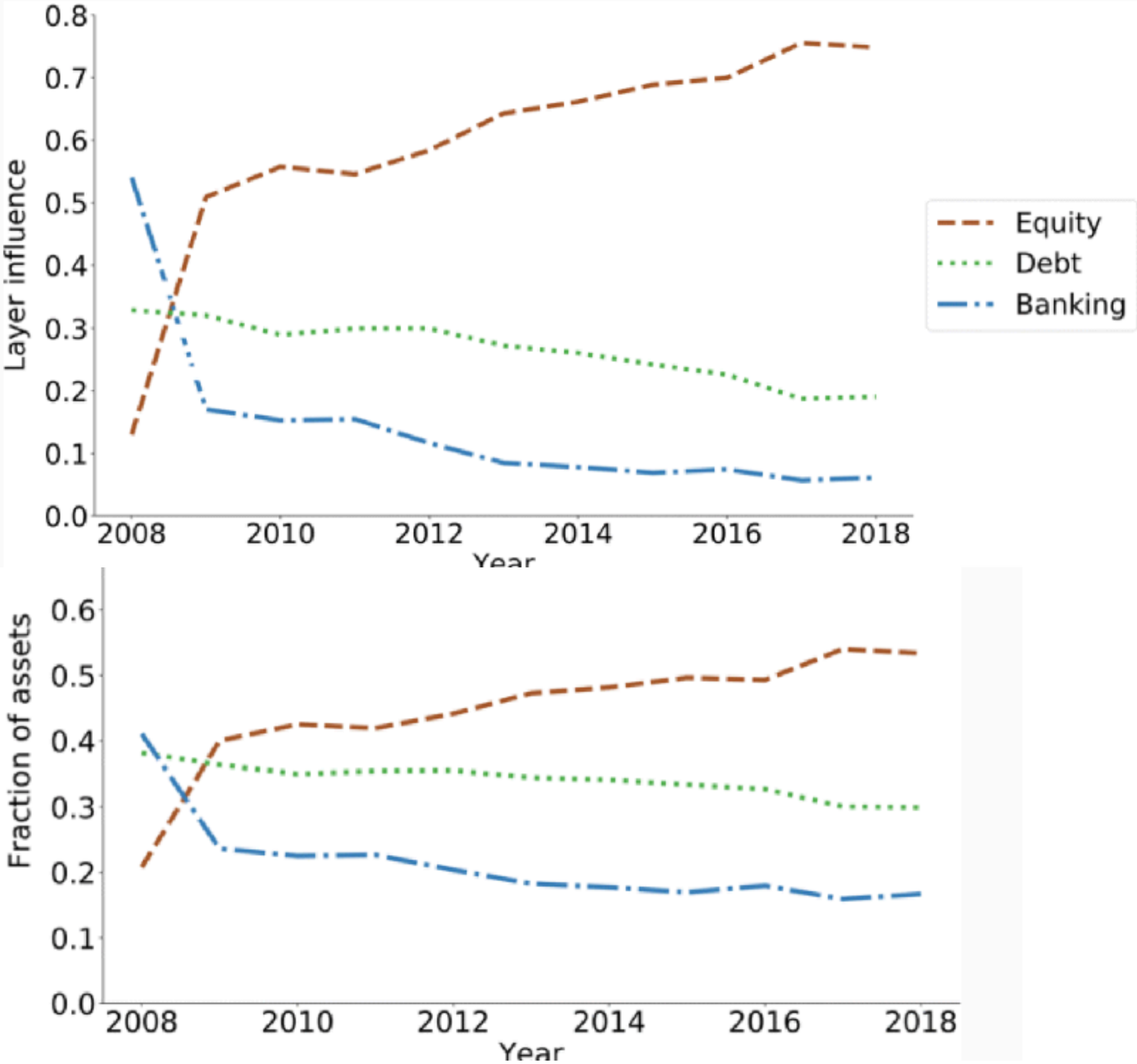


M. DEL RIO-CHANONA ET AL. **“THE MULTIPLEX NATURE OF GLOBAL FINANCIAL CONTAGIONS”**
https://appliednetsci.springeropen.com/articles/10.1007/s41109-020-00301-2

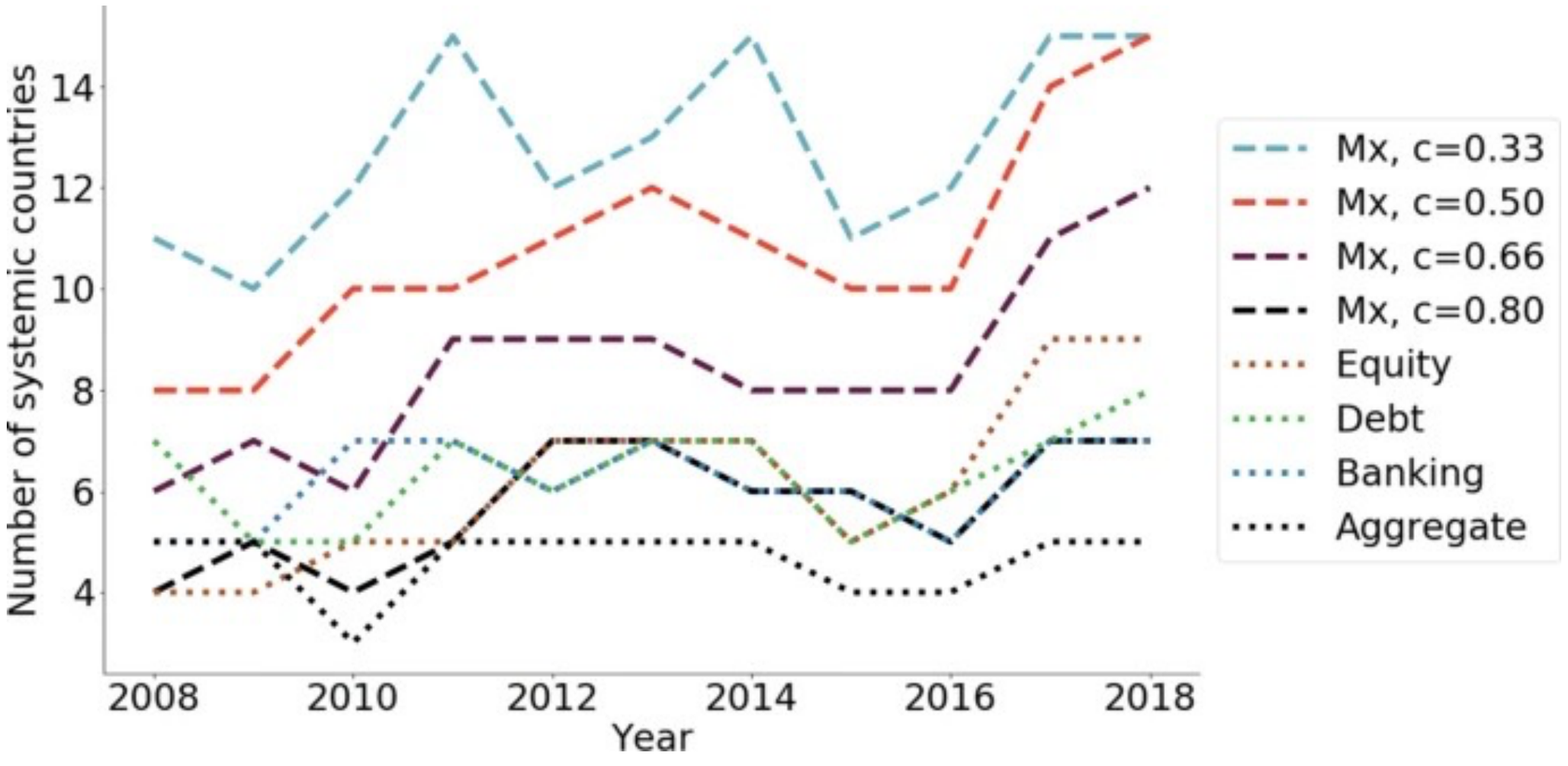
CASE STUDY II – FINANCIAL CONTAGION



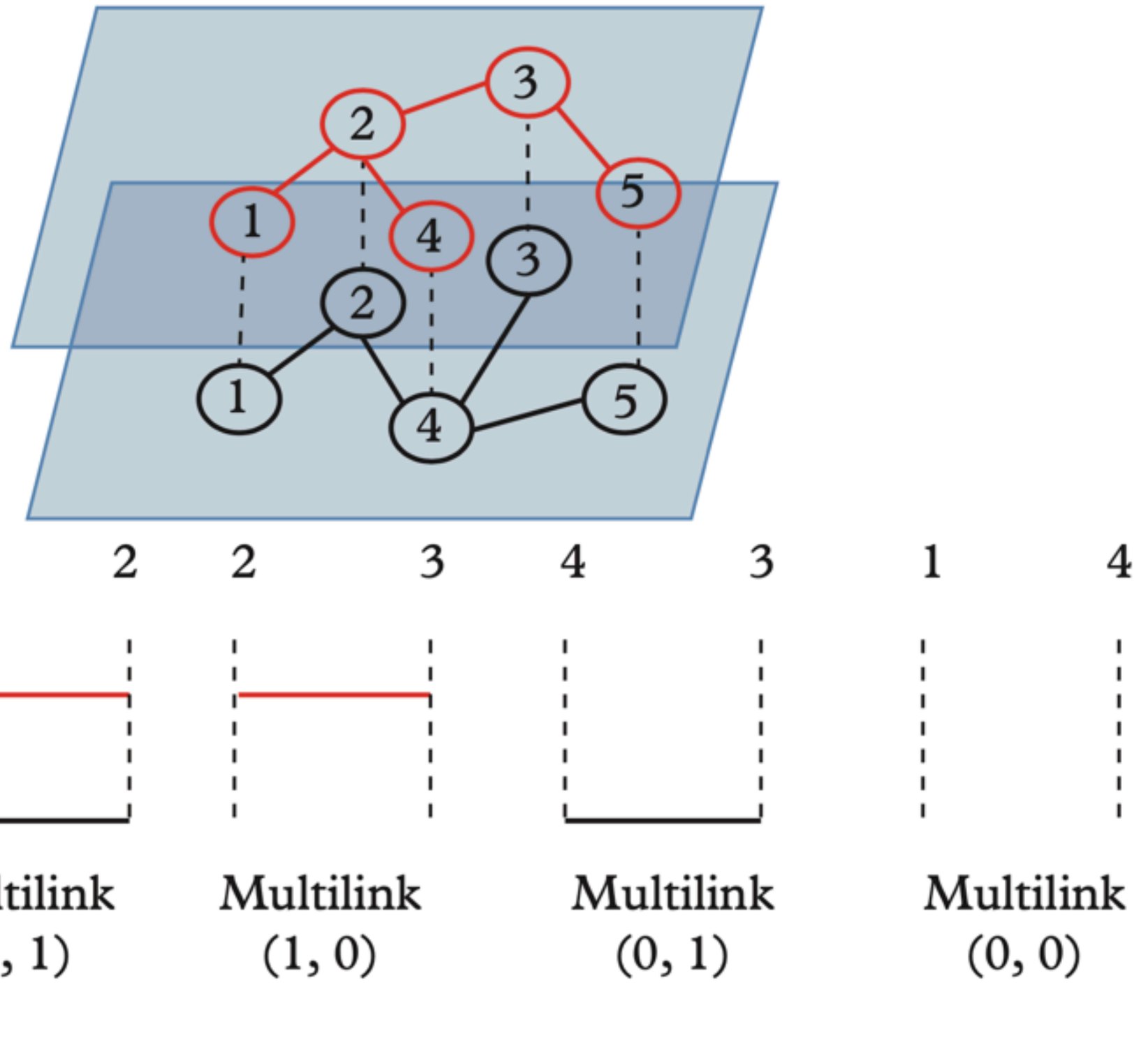
M. DEL RIO-CHANONA ET AL. **"THE MULTIPLEX NATURE OF GLOBAL FINANCIAL CONTAGIONS"**
https://appliednetsci.springeropen.com/articles/10.1007/s41109-020-00301-2


Layer influence
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
2008
2010
2012
2014
2016
2018
Year
Equity
Debt
Banking
Fraction of assets
0.0
0.1
0.2
0.3
0.4
0.5
0.6
2008
2010
2012
2014
2016
2018
Year

A single-layer approach underestimates risk



CASE STUDY II.b – Multiplex trees



Multilinks

$$z^{(1,0)} = \sin\theta\cos\phi,$$

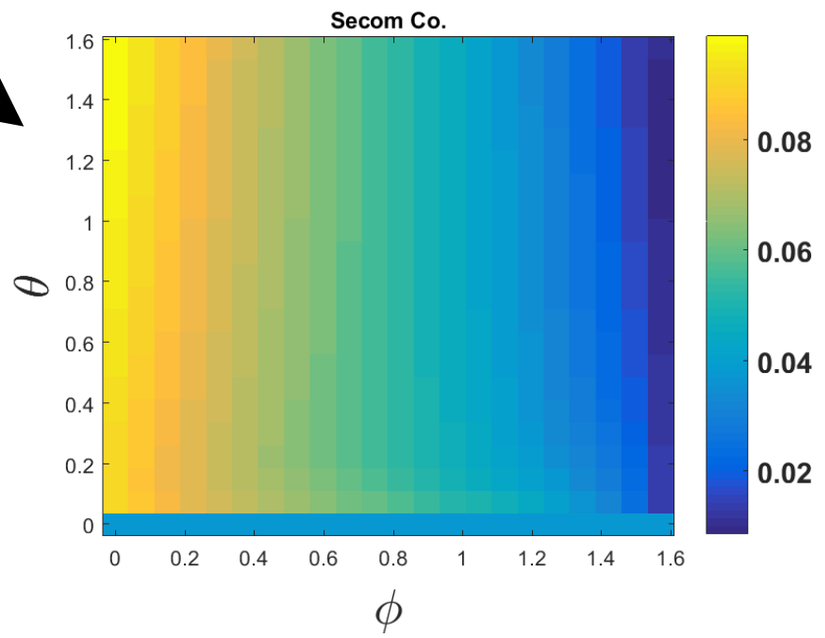
$$z^{(0,1)} = \sin\theta\sin\phi,$$

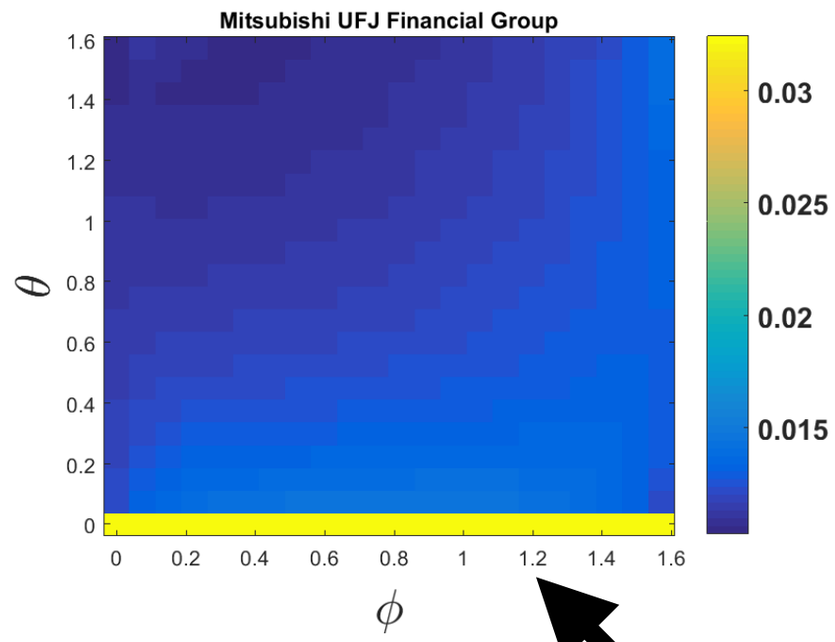
$$z^{(1,1)} = \cos\theta,$$

$$\theta, \phi \in [0, \pi/2]$$

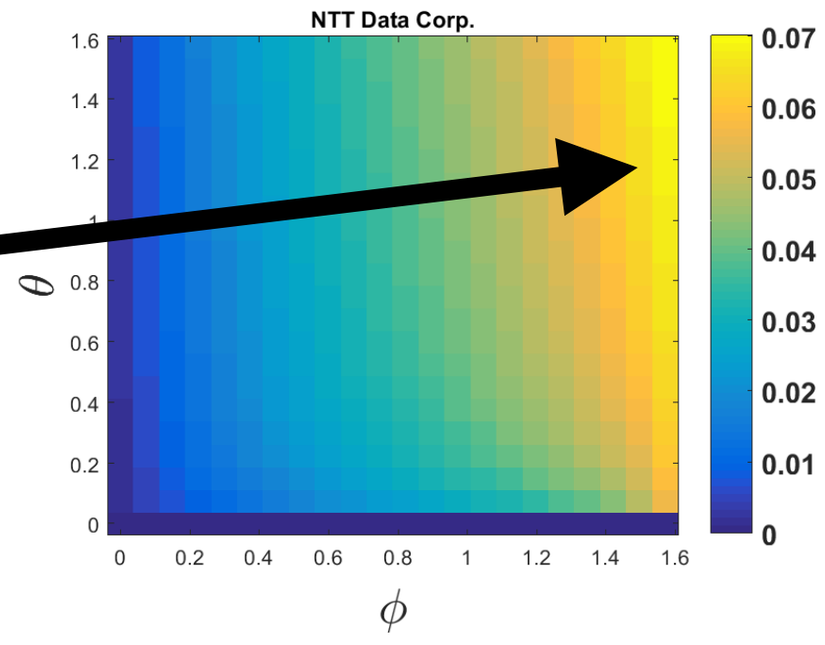
Centrality maps

Price





Volume



Both layers Only