Patent war dynamics
Week 2.
Data filterting and network construction

...

Valo Hallman, Jenny-Rosa Mattila, Stanislav Andreev

What does the data look like

~100 000 cases between 2000-2021

case_node_id	Case Title	Civil Action #	Venue	Filing Date	included in random sample	DJ	patents	Alleged Infringer	paragraph_id	Patent Asserter	asserter_id	Asserter Category
118525	Williams, et al v. Ger	1:00-cv-00037	E.D.Tex.	2000-01-13	1	0	5655545	Tyco International, L	4015	Randy Williams	97476	9
118525	Williams, et al v. Ger	1:00-cv-00037	E.D.Tex.	2000-01-13	1	0	5655545	Tyco International, L	4016	Lana Davis Johnso	80621	9

Patent Asserter
Randy Williams
Lana Davis Johnso

Alleged Infringer
Tyco International, L
Tyco International, L

patents
5655545
5655545

Who Sues who on what

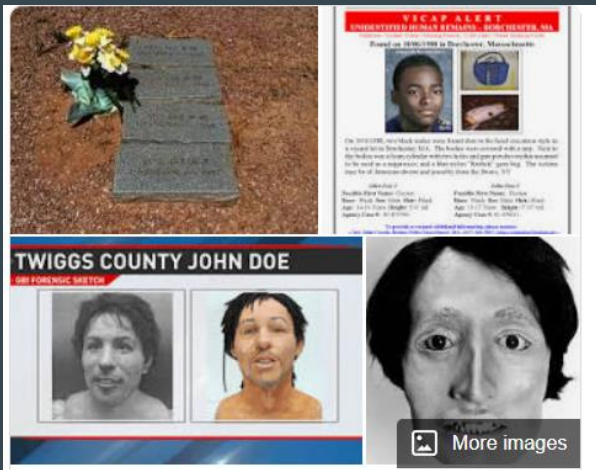
Problems with the data

- Different names of the same company
- Human errors (typos, different format, abbreviations)

```
Len of unique alleged infringers before filtering: 74728
```

```
[('Does', 1653),
 ('Teva Pharmaceuticals USA, Inc.', 1106),
 ('Mylan Pharmaceuticals, Inc.', 1100),
 ('Apotex Inc.', 893),
 ('Apple Inc.', 810),
 ('Mylan, Inc.', 708),
 ('Samsung Electronics America, Inc.', 701),
 ('LG Electronics, Inc.', 663),
 ('Sandoz Inc.', 663),
 ('Samsung Electronics Co., Ltd.', 634),
 ('Lupin Ltd.', 612),
 ('Amazon.com, Inc.', 605),
 ('Teva Pharmaceutical Industries Ltd.', 592),
 ('Microsoft Corporation', 576),
 ('Lupin Pharmaceuticals, Inc.', 572)]
```

```
toyota': ['Toyota Motor Corporation', 'Toyota Motor Sales, U.S.A., Inc.'],
motorola': ['Motorola Mobility, Inc.',
'Motorola Solutions, Inc.',
'Motorola Mobility LLC',
'Motorola, Inc.',
'Motorola Solutions, Inc.',
'Motorola Mobility LLC',
'Motorola, Inc.',
'Motorola Mobility, Inc.',
'Motorola Mobility LLC',
'Motorola, Inc.',
'Motorola Mobility, Inc.',
'Motorola Solutions, Inc.'],
```



John Doe

John Doe and Jane Doe are multiple-use placeholder names that are used in the United States and the United Kingdom when the true name of a person is unknown or is being intentionally concealed. Wikipedia

Merge possible duplicates (asserters and alleged infringers)

- Filtering algorithm

- 1. Find the most frequent terms
- 2. Filter out technical parts (like
- 3. Make pairwise distance between terms
  python)
- 4. Merge connected (distance < threshold) terms and assign the same alias for all of them.
- 5. Iterate over all terms and find matches with aliases from 4.
- 6. Filter out all asserters and alleged infringers with number of cases less than threshold (e.g. 3)

```
ignore_words_ = ["Corporation", "Inc.", "Ltd.", "LLC", "Co.", "Co", "Electronics", "Technologies", "Industries", "Group", "Holdings", "Companies", "Company",
                 "Device", "Digital", "International", "Global", "Telecommunications", "Semiconductor", "Semiconductors",
                 "Systems", "Solutions", "International", "America", "USA", "US", "United States", "Company", "Motor", "Sales", "Manufacturing", "Manufacture", "Manufactures",
                 "North America", "South America", "Europe", "Asia", "Africa", "Australia", "Pacific", "U.S.A.", "Mobility", "LLC",
                 "Communications", "Laboratories", "Pharmaceutical", "Pharmaceuticals", "Pharma", "Biotech", "Biotechnologies", "corp.", "partnership",
                 "incorporated", "of", "the", "and", "in", "for", "on", "with", "by", "at", "to", "from", "as", "an", "or", "is", "are", "was", "were", "be", "been", "being",
                 "services", "Microelectronics", "North", "Electronic", "wireless", "americas",
                 "stores", "direct", "sprint", "network", "networks", "mobile", "telecom", "telecoms", "telecomms", "telecommunications", "telecomunication", "telecomunications",
                 "Limited", "Healthcare", "Health", "Medical", "Medicine", "Medicines", "Therapeutics", "Therapy", "Therapies", "Computer", "Software", "Inc", "ltd", "united", "states", "pr
```

[jupyter notebook with the code]

```
Len of unique alleged infringers before filtering: 74728
```

```
Number of infringers with count > 3: 9601
```

```
Number of patent asserters with count > 3: 5081
```

Construct network

- Node – company/individual (alias from map produced in the previous step)
- Directed edge reflects asserted case. Weight – number of cases.
- We use only active asserters/infringers (filtering by threshold)

To track dynamics over years we splited data into chunks related to 3 year periods.

```
'samsung': ['Samsung Electronics Co., Ltd.',
 '"Samsung"',
 'Samsung America, Inc.',
 'Samsung SDI Co., Ltd.',
 'Samsung SDI America, Inc.',
 'Samsung Electronics America, Inc.',
 'Samsung Telecommunications America LLP',
 'Samsung Semiconductor, Inc.',
 'Samsung Telecommnications America, L.P.,',
 'Samsung Austin Semiconductor, L.P.',
 'Samsung Information Systems America, Inc.',
 'Samsung Telecommunications America General, L.L.C.',
 'Samsung Austin Semiconductor, LLC',
 'Samsung Electro-Mechanics America, Inc.',
 'Samsung Electro-Mechanics Co., Ltd.',
 'Samsung International, Inc.',
 'Samsung Electronics Canada Inc.',
 'Samsung Telecom America LLP',
 'Samsung Technology, Inc.',
 'Samsung Telecommunications America LLC',
 'Samsung Telecommunications America, Inc.',
 'Samsung America',
 'Samsung Opto-Electronics America, Inc.',
 'Samsung Techwin Co.,',
 'SAMSUNG AUSTIN GENERAL L.L.C.',
 'Samsung Electronics American, Inc.',
 'Samsung Elec. Co. Ltd.',
 'Samsung Electronics America, L.P.',
 'Samsung Techwin Co., Ltd.',
 'Samsung Semiconductors, Inc.',
 'Samsung Elecronics Co., Ltd.',
 'Samsung Electronics America, LLC',
 'Samsung Mobile Display Co., Ltd.',
 'Samsung Semiconductor Europe GMBH',
 'Samsung Electronics Research Institute',
 'Samsung Networks America, Inc.',
 'Samsung Digital Imaging Co., Ltd.',
 'Samsung Telecommunicatons America LLC',
 'Sansung Electro Mechanics America Inc',
 'Samson Technologies Corp.',
 'Samyung Enc Co. Ltd.',
 'Samsung Electronics Co.',
```

Visualization

To simplify the plot we showed only nodes with degree more than some threshold (it was increased for later years)

Reciprocated edges are especially interesting for us (if companies made cases against each other), so we highlighted them in red color.

To reflect 2 possible sides of patent war we used **spring_layout** from networkx, which utilizes Fruchterman-Reingold force-directed algorithm. Nodes imitated repelling objects (called an anti-gravity force)*. In our model central nodes are usually produce the cases against border nodes.

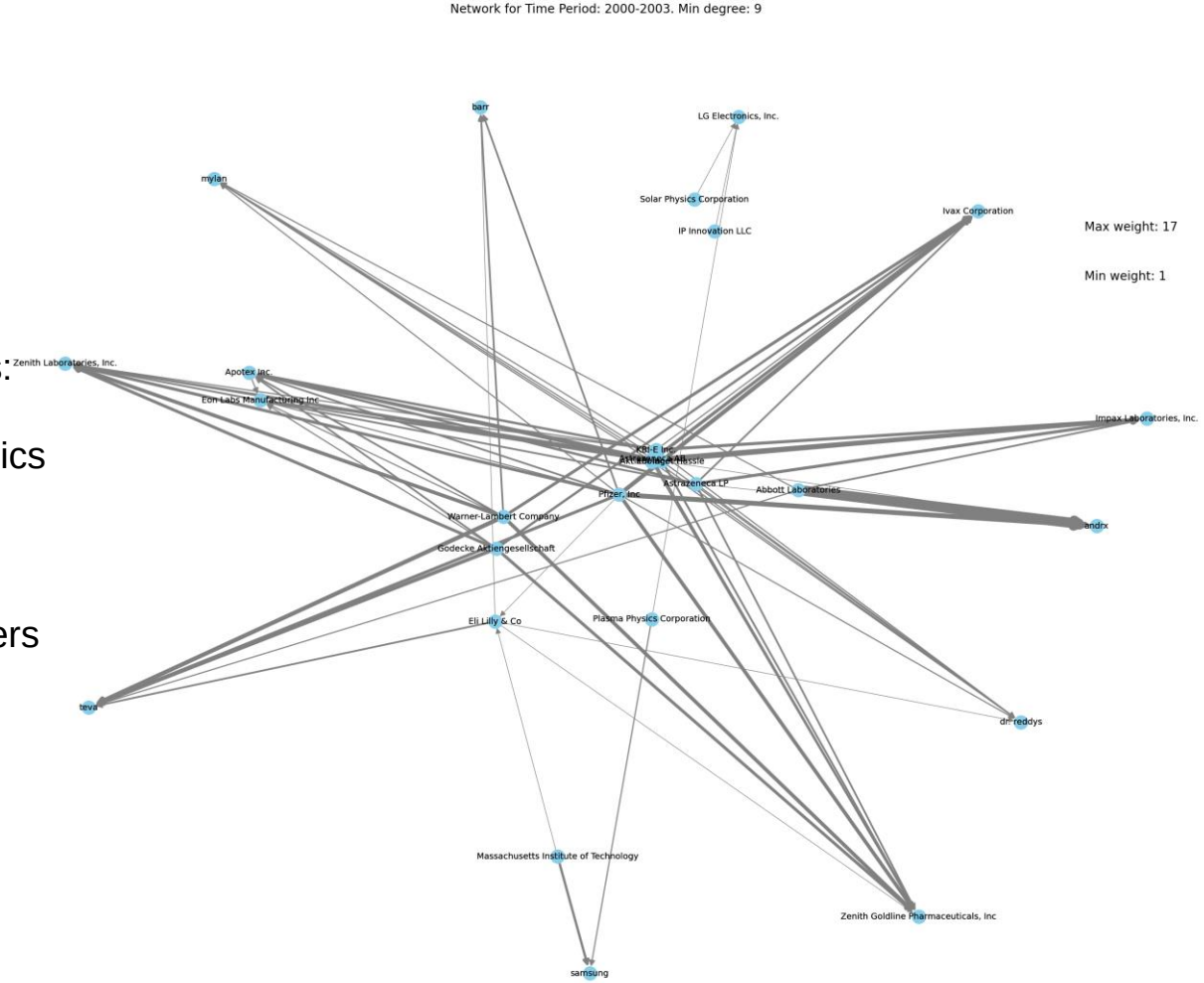
https://networkx.org/documentation/stable/reference/generated/networkx.drawing.layout.spring_layout.html

Visualization

We can distinguish 2 clusters:

- pharma
- small technology/electronics cluster

- Some Institutes/law companies connect clusters



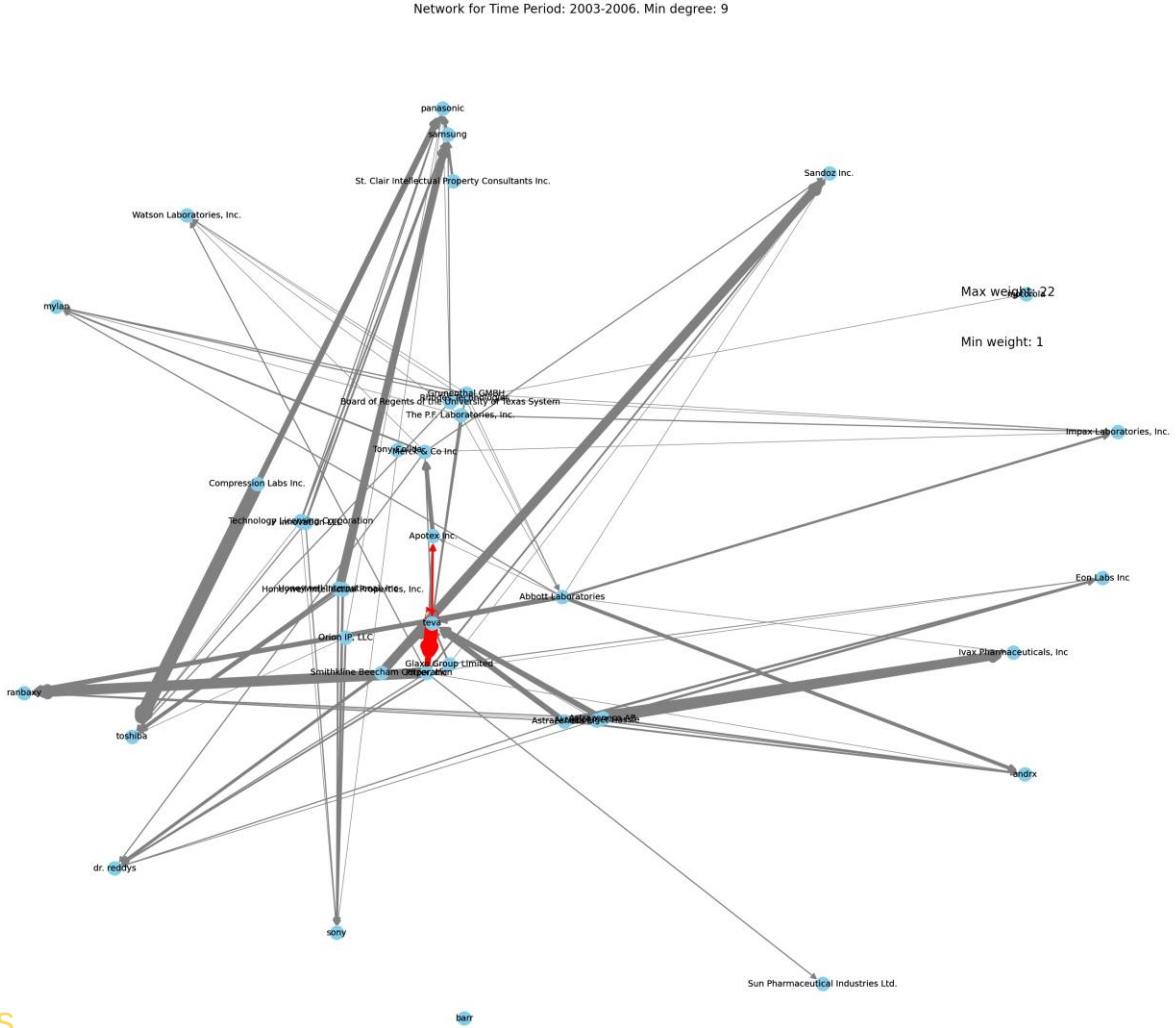
Jupyter notebook with the code and plots

Visualization

Still the same clusters

+ counterclaims

Relatively the same volume

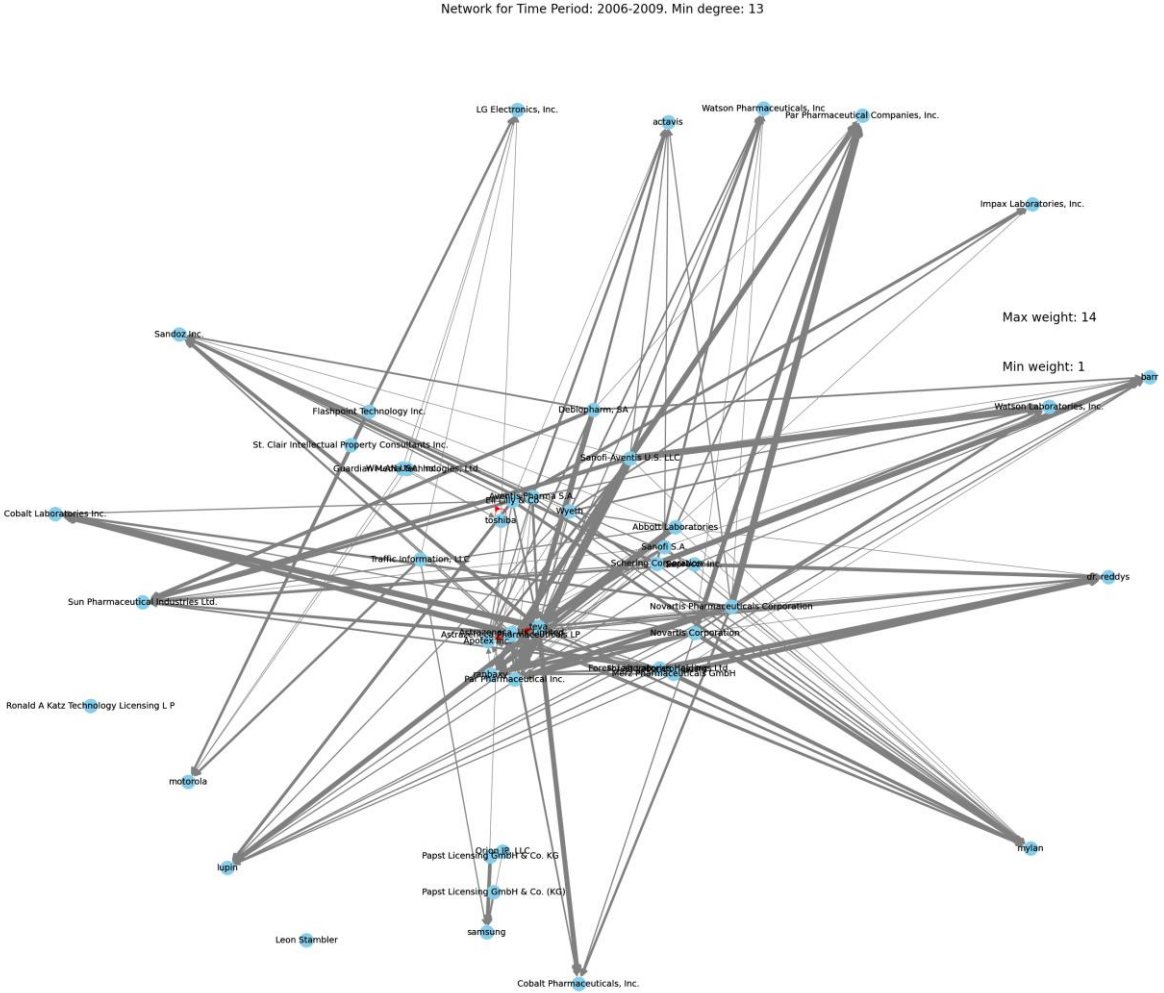


[Jupyter notebook with the code and plots]

Visualization

Bigger volume, we increased threshold for min degree

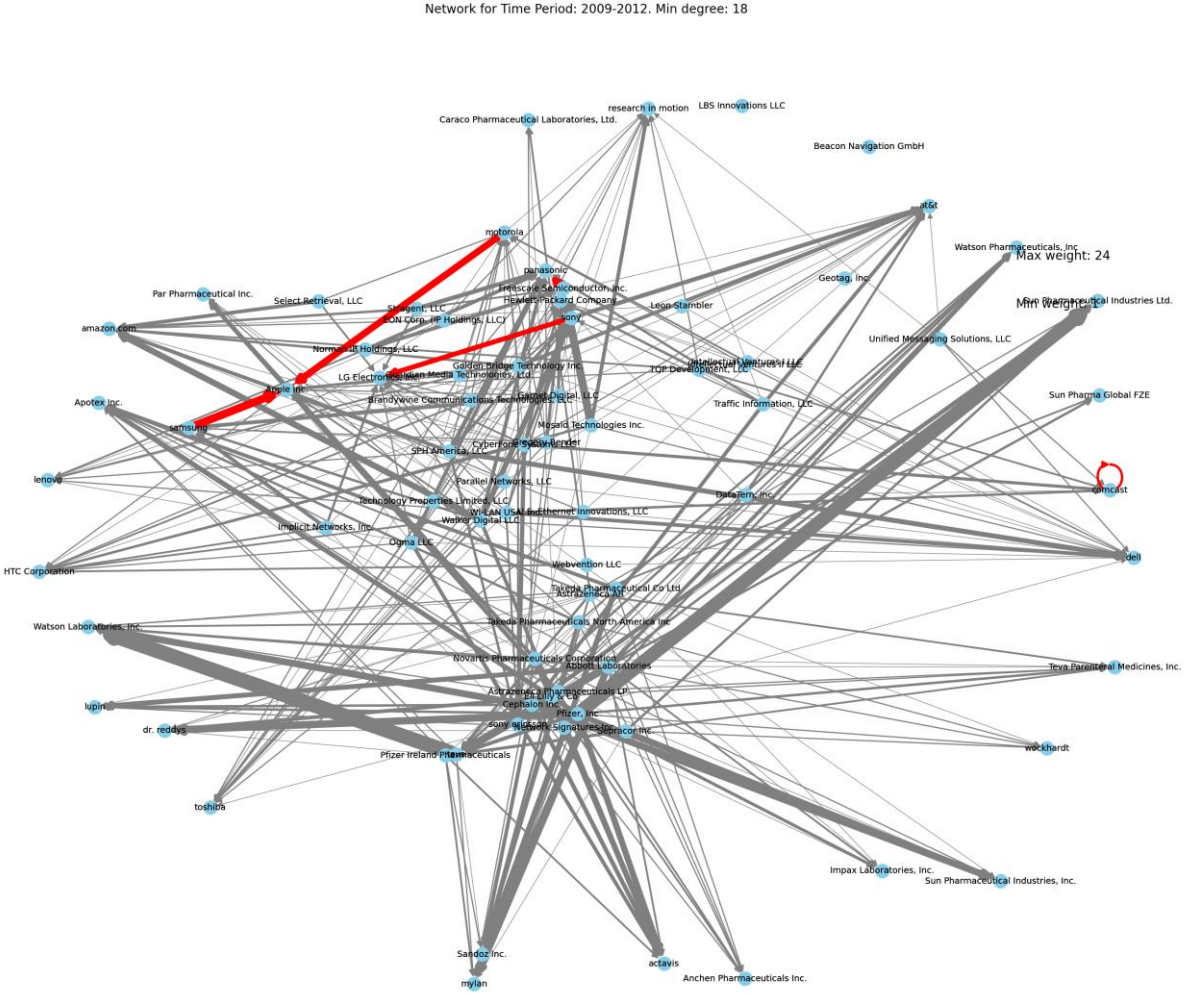
More licensing companies



Jupyter notebook with the code and plots

Visualization

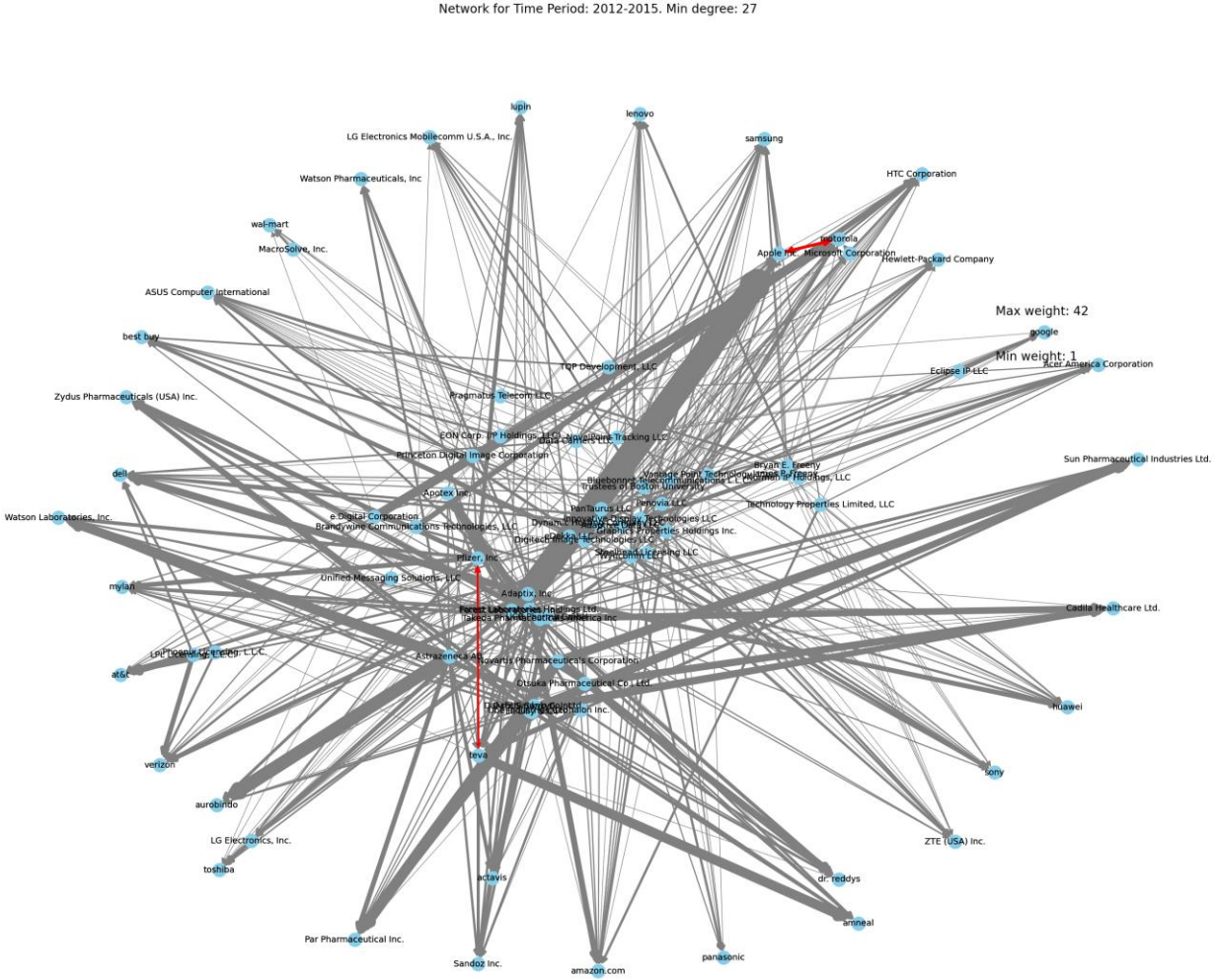
- Counterclaims between technology giants (Apple, Samsung, Motorola, LG, Sony)
- Pharma and technology clusters are recognizable



Jupyter notebook with the code and plots

Visualization

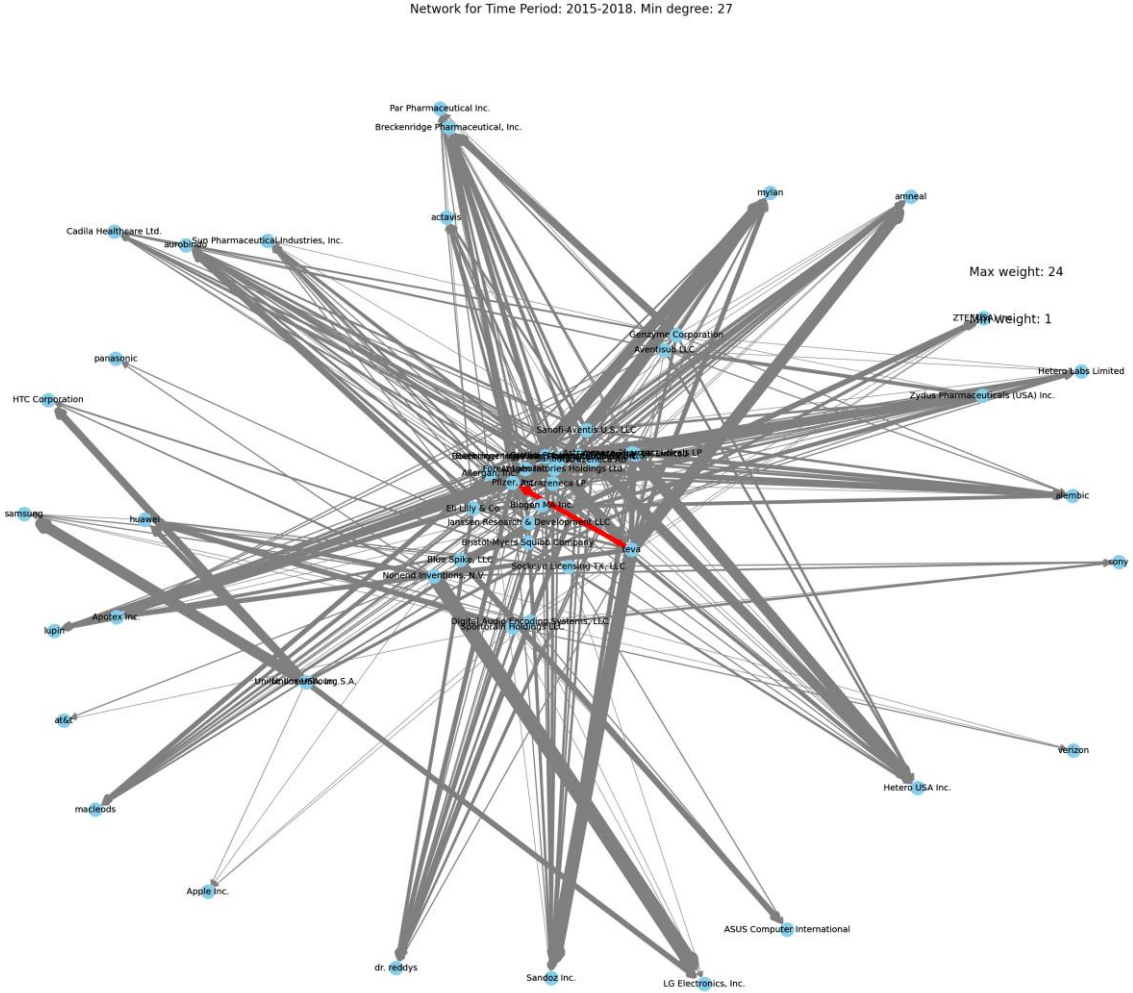
- Less counterclaims between technology giants (Apple vs Motorola)
- Migration of companies from center (attackers) to borders and vice versa
- Dramatic increase of cases volume (min degree 27)



[Jupyter notebook with the code and plots]

Visualization

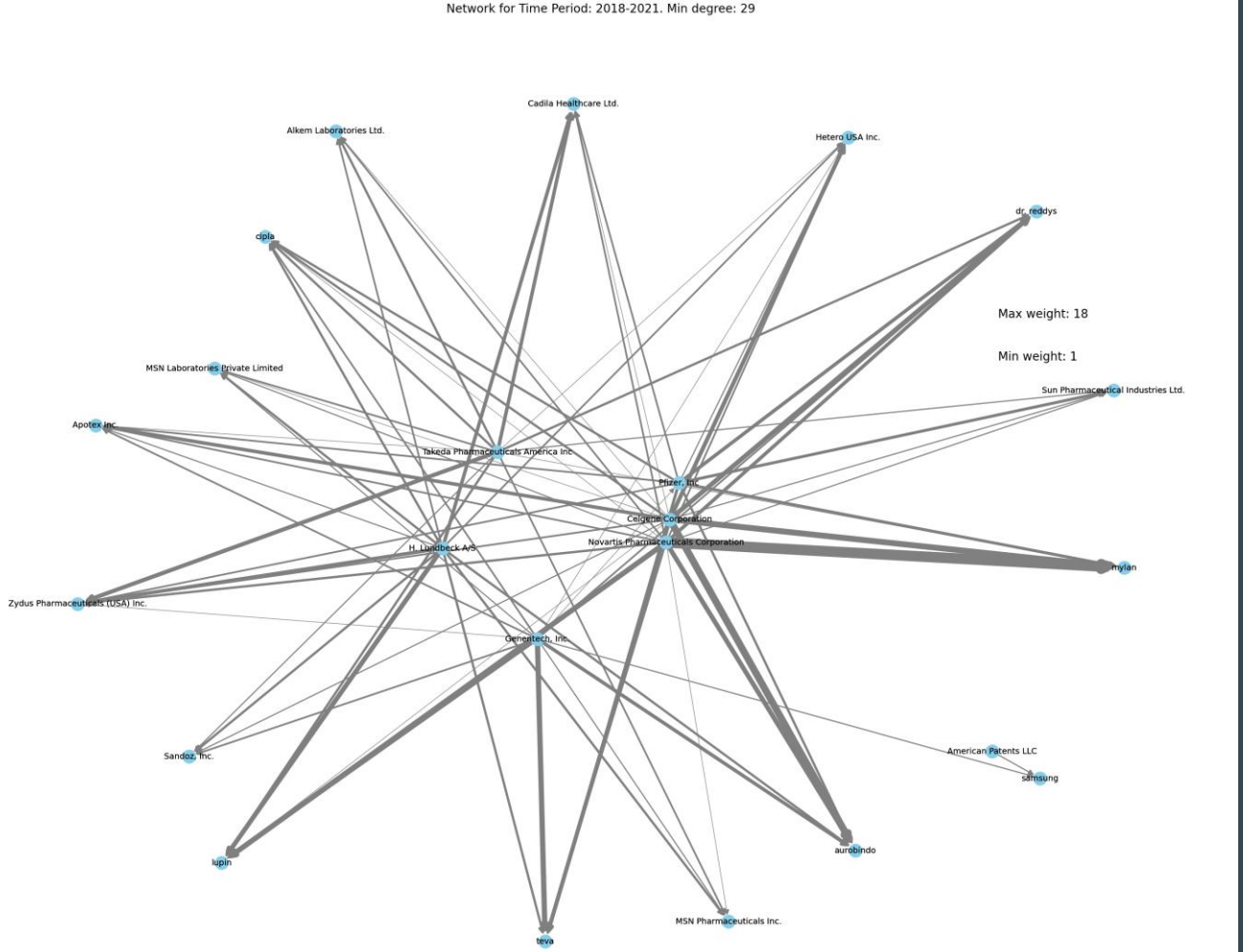
- No counterclaims between technology giants



[Jupyter notebook with the code and plots]

Visualization

- Pfizer still in the center
- It can indicate, that in pharma and technology clusters successful market position can lead to different places in such graph representation

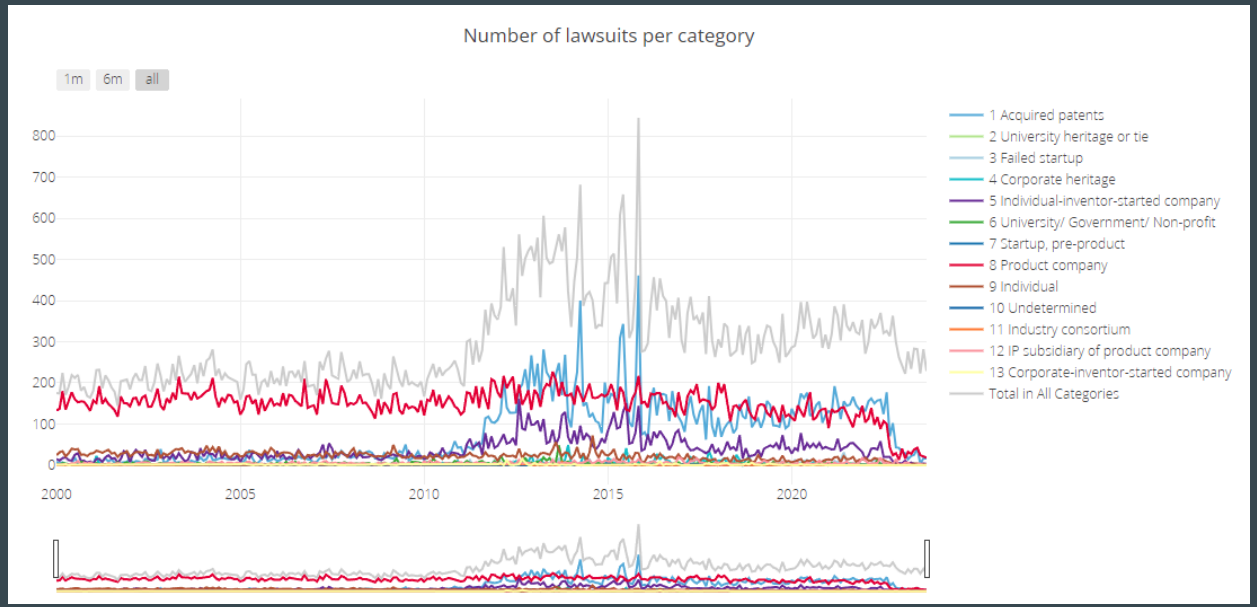


Jupyter notebook with the code

What next

-Investigate lawsuits categories (presented in data)

-Try dynamic visualization



https://npe.law.stanford.edu/

Thanks!