Submission for PhD position in computational social sciences at ETH Zürich

March 2, 2020

Marius Herget

Contents

Task 1

My first experience with research was within my practical training for my B.Sc. employer at the IBM Almaden Research Center in 2016. I experienced first hand the enthusiasm of the scientist in asking and pursuing groundbreaking questions. After these three months, I decided to follow my inner craving to change the world. Especially the scientific approach with no prejudices, religion, or political agenda matches my inner conviction of just listening to facts. These experiences convinced me to achieve the highest university degree. Within my life so far, I always set myself the standard to obtain and deliver the best possible outcome. Consequently, I would like to continue my career at one of the best universities in Europe.

My ambitions for my Ph.D. position is a mix of a great team and flexibility. Exploring new areas, receiving founded advice, and the willingness to surpass oneself should be the daily business. Nevertheless, the essential aspect of any future career step is the working environment. The personal and professional climate has to be friendly and giving me a strong impetus. Research wise I am very open-minded where the journey goes. I love the challenge of familiarizing myself with new facts and theories.

Task 2

As already mentioned in my motivation letter, I think my technical skills are a perfect match for your team. The most exciting aspect is the practical implementation of various technologies (especially computational science) on social and political systems. These complex, traditional systems have a massive impact on society, and a critical analysis is crucial to regulate them.

Nevertheless, your chair is accomplishing quite exciting research in more detail *multi-layered networks* and *socio-technical systems*. The aspect of the data-driven analysis is an innovative approach with great potential. In more detail, the paper *"Quantifying and suppressing ranking bias in a large citation network"* [3] has aroused my interests. The recognized biases within papers ranking and, therefore, their "false" importance is an enormous disruption of an old standing tradition of the scientific world.

Task 3

Since the dataset is limited to only one specific research field, I would like to ask the following research question as a base for my data analysis:

Exist unnoticed differences in wording within various research groups of a similar research area?

My analysis would take several steps:
1. Identify research groups based on the connection of *author*, *affiliations* with references of their publication to other papers and their authors (*citations*).
2. Identify significant words within the *title*, *abstract*, and *full text*. This includes a data cleanup of common words like *is*, *are*, *they*, *and*, etc.
3. Connect both results of steps 1. and 2. to recognize which research group applies which wording (= wording bubbles).
4. Classify word bubbles and review their significance.
5. Examine significant words by connecting synonyms within the dataset. Compare wording bubbles and identify via step 5. how various research groups differ in phrasing and content.

The goal of this analysis is to recognize whether scientists miss relevant research publications in consequence of their **wording filter bubble**.

This could be quite interesting for identifying other important papers while minimizing research expenses. On the one hand, in a practical application, a reduction of reading the current state of the art can be quite impelling. On the other hand, a broadening of the horizon can enhance a publication's quality.

Task 4

This task is centered around the fact of finding the most important member of the House. Firstly it is crucial to define what importance means in this context:

An important person is an individual who has the most power within the House. There are several aspects to consider:

Money An influential person needs financial stability and has a high budget.

Reelections Besides financial stability, a politician depends on its position in his party and its district's population (voters). Another aspect is that a member of the House gains more importance with his experience and the broader network.

Bills Another aspect besides the politicians standing is the importance of his sponsored bills. In this assignment, this is shrunk down to the number of cosponsors and their party. It is more challenging to receive cosponsors from the opposite party.

This is only a small sample of a lot of different aspects to recognize the most important person.

The following pseudo equations give a quick overview how the importance is calculated

$$Imp_{Bill} = (1.1 * \frac{CS_{otherParty}}{CS_{all}} + 0.9 * \frac{CS_{ownParty}}{CS_{All}}) * Normalized(CS_{Total}) \tag{1}$$

$$Imp_{Representative} = \frac{1}{2} * Normalized(\sum Money_{Individual}) + \frac{1}{2} * \frac{Periods\ in\ House_{last\ 4}}{4} \tag{2}$$

$$Importance = \frac{1}{2} * Imp_{Representative} + \frac{1}{2} * Normalized(\sum_{Sponsor=Individual} Imp_{Bill}) \tag{3}$$

$$\tag{4}$$

To achieve the correct analysis, the given data had to be supplemented by [1] and [2].

Edward R. Royce is according to my analysis with a significant lead the most import member:

Party Republican
Bills in 115 33
Available Money 7340372.91$
In House of Representatives since 1993
Received cosponsorships 1685
Given cosponsorships 212
Cosponsorships from Democrats 822
Cosponsorships from Republicans 864
Importance$_{Representative}$ 0.713342
Importance$_{Bill}$ 4.292423
Importance 0.856671

References

[1] *Bills - ProPublica Congress API.* [Online; accessed 24. Feb. 2020]. Jan. 2020. URL: `https://projects.propublica.org/api-docs/congress-api/bills/#get-cosponsors-for-a-specific-bill`.

[2] *House Office Expenditure Data.* [Online; accessed 24. Feb. 2020]. June 2019. URL: `https://projects.propublica.org/represent/expenditures`.

[3] Giacomo Vaccario et al. "Quantifying and suppressing ranking bias in a large citation network". In: *Journal of Informetrics* 11.3 (Aug. 2017), pp. 766–782. ISSN: 1751-1577. DOI: `10.1016/j.joi.2017.05.014`.

Appendices

Table 1: Top ten most important members

Name	Party	Bills	Received	Given	Money	Reelection	Imp_{Rep}	CS_D	CS_R	Imp_{Bill}	Imp
Royce, Edward R.	Republican	33	1685	212	7340372.91	1.00	0.713342	822	864	4.292423	0.856671
Cicilline, David N.	Democrat	50	1689	674	10138790.22	0.75	0.670829	1641	47	3.862105	0.785289
Maloney, Carolyn B.	Democrat	41	1329	435	11664144.51	1.00	0.840792	1195	136	3.092045	0.780571
DeLauro, Rosa L.	Democrat	38	1338	409	10629422.22	1.00	0.810292	1307	31	3.061348	0.761745
Engel, Eliot L.	Democrat	42	1289	500	8936514.00	1.00	0.760391	956	333	3.090439	0.740183
Paulsen, Erik	Republican	39	1328	203	7100318.85	1.00	0.706266	563	766	3.305533	0.738176
Lee, Barbara	Democrat	45	1217	751	10749907.02	1.00	0.813843	1178	39	2.765735	0.729086
Velazquez, Nydia M.	Democrat	43	1129	584	10383839.10	1.00	0.803053	949	186	2.653293	0.710593
Reichert, David G.	Republican	22	1161	145	7412045.22	1.00	0.715455	655	515	3.024453	0.710029
Roe, David P.	Republican	30	1140	309	9508300.44	1.00	0.777245	268	879	2.752337	0.709227

Imp: Importance, CS: Cosponsors, Rep: Representative, D: Democrat, R: Republican, Money in $