Appendix

1 Distributions of motifs

1.1 Computing the distribution probabilities of motifs

Let us define the *multi-set* $\mathcal{S}$ to contain all the trajectories present in the data. For example, given a trajectory $i = \{A \to B \to A \to B\}$ appearing n_i times in the data, then i appears n_i times in $\mathcal{S}$. For each trajectory $i \in \mathcal{S}$, we also define the *multi-set* $\mathcal{S}_k(i)$ to contain all the sub-trajectories of i with length k. For example, given a trajectory $i = \{A \to B \to A \to B\}$, the *multi-set* $\mathcal{S}_1(i) = \{A \to B, A \to B, B \to A\}$. Note that the sub-trajectory $A \to B$ appears twice in i, i.e., $m(A \to B, i) = 2$, while $B \to A$ appears only once, i.e., $m(B \to A, i) = 1$. In general, we have that $|\mathcal{S}_k(i)| = \max(l_i - k + 1, 0)$ where l_i is the length of trajectory i. In other words, a trajectories of length l_i can be split into $l_i - k + 1$ sub-string of length k if $k \leq l_i$. This can also be expressed as $\sum_{p \in \tilde{\mathcal{S}}_k(i)} m(p, i) = \max(l_i - k + 1, 0)$ where $\tilde{\mathcal{S}}_k(i)$ is the *set* of sub-trajectories of length k extracted from i.

To compute the probability to observe motifs of type I, i.e., $X \to Y \to X$ or type II, i.e., $X \to Y \to Z$, we have note that these are all the possible (sub-)trajectories of length two. We call $\mathcal{S}_k$ the *multi-set* containing all the sub-trajectory of length k. This means that $\mathcal{S}_2$ contains all the motifs of type I and type II. Then, we define the **probability to observe a motif** $p \in \mathcal{S}_2$

$$P^{(\mathrm{emp})}(p) = \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} q(p|i) = \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} \frac{m(p, i)}{\max(l_i - 1, 1)} \tag{1}$$

Note that $P^{(\mathrm{emp})}$ is is always equal or greater than zero and $\sum_{p \in \tilde{\mathcal{S}}_2} P^{(\mathrm{emp})}(p) + |\mathcal{S}_1|/|\mathcal{S}| = 1$ where $\tilde{\mathcal{S}}_2$ is the *set* containing the sub-trajectory of length two. This last relation follows from the fact that any trajectory $i \in \mathcal{S}$ can be spitted in $l_i - 1$ sub-trajectories of length 2 when $l_i \geq 2$. Hence,

$$\sum_{p \in \tilde{\mathcal{S}}_2} q(p|i) = \begin{cases} 1 & \forall i \,:\, l_i \geq 2 \\ 0 & \forall i \,:\, \text{otherwise} \end{cases} \tag{2}$$

Using Eq.2, we can show that

$$\sum_{p \in \tilde{\mathcal{S}}_2} P^{(\mathrm{emp})}(p) = \frac{1}{|\mathcal{S}|} \sum_{i \in \tilde{\mathcal{S}}} n_i \sum_{p \in \tilde{\mathcal{S}}_2} q(p|i) = \frac{1}{|\mathcal{S}|} \sum_{k=2} |\mathcal{S}_k| \Longrightarrow \sum_{p \in \tilde{\mathcal{S}}_2} P^{(\mathrm{emp})}(p) + \frac{|\mathcal{S}_1|}{|\mathcal{S}|} = 1 \tag{3}$$

For computing the probability distribution in the first- and second-order network, i.e., $P^{(1)}$ and $P^{(2)}$, we rely on the `Pathpy` implementation. Precisely, we use the function `path_likelihood` that returns the probability to observe a path/transition. For more details, see (**??**).

1.2 Odds ratio

We split the set $\tilde{\mathcal{S}}_2$ in two disjoint sets $\tilde{\mathcal{S}}_2^I$ and $\tilde{\mathcal{S}}_2^{II}$,respectively containing motifs of type I and type II. The empirical odds ratio to observe motifs of type I compared to motifs of type II is given by

$$OR_{I,II} = \frac{\sum_{p \in \tilde{\mathcal{S}}_2^I} P^{(\text{emp})}(p)}{\sum_{p \in \tilde{\mathcal{S}}_2^{II}} P^{(\text{emp})}(p)} \tag{4}$$

and we find from our data that $OR_{I,II} \sim 1.7$. This means that motifs of type I are almost twice as frequent as motifs of type II. When computing the odds ratio $OR_{I,II}$ using $P^{(1)}$ we get ~ 0.14 that is of an order magnitude different compared to the empirical one. While, the the odds ratio $OR_{I,II}$ coming from $P^{(2)}$ is ~ 1.4 that is quite close to the empirical one.

1.3 Computing and visualizing the Kullback-Leibler divergence

Since we use the $P^{(\text{emp})}$, $P^{(1)}$, $P^{(2)}$ to compute the Kullback-Leibler (KL) divergence for motifs of length 2, we normalize this probability to one. In other words, for $P^{(\text{emp})}(p)$, we write

$$P^{(\text{emp})}(p) \to P^{(\text{emp})}(p) / \sum_{p \in \tilde{\mathcal{S}}_2} P^{(\text{emp})}(p) \tag{5}$$

and we renormalise in similar way $P^{(1)}$ and $P^{(2)}$. The KL-divergence between $P^{(\text{emp})}$ and $P^{(1)}$ is 1.51, while between $P^{(\text{emp})}$ and $P^{(2)}$ is 0.10. Hence, the first-order model has KL-divergence more than 10 times larger compared to the second-order model. In Fig. 1, we plot $P^{(\text{emp})}$, $P^{(1)}$, and $P^{(2)}$ respectively in green dots, blue + and orange $\times$. The order on the x-axis is created by sorting the motifs from the most to least probable with respect to the empirical data. We see that the first-order network consistently underestimates motifs that are more probable. Instead, the second-order network produces probabilities quite close to the empirical ones.

To compute the Kullback-Leibler divergence only for motifs of type I (type II), we have to renormalise $P^{(\text{emp})}$, $P^{(1)}$, and $P^{(2)}$. In other words, $P^{(\text{emp})}(p) \to P^{(\text{emp})}(p) / \sum_{p \in \tilde{\mathcal{S}}_2^I} P^{(\text{emp})}(p)$, and similarly for $P^{(1)}$ and $P^{(2)}$. In Fig. 2(a) and (b), we compare the probability distribution for the motifs of type I and type II, respectively. Again, we have $P^{(\text{emp})}$, $P^{(1)}$, and $P^{(2)}$ respectively in green dots, blue + and orange $\times$. The order on the x-axis is created by sorting the motifs from the most to least probable with respect to the empirical data. From Fig. 2(a) that depict motifs of type I, it is evident that the first order network underestimates the probabilities of these motifs. While, in Fig. 2(b), the situation is reversed.

2 Trajectories at city level

In MEDLINE, we have 3 740 187 individual scientist trajectories across 12 980 cities between 1990 and 2009.. Among these, 884 251 trajectories have length 1 or higher. Specifically, 11 % of all the trajectories are of length 1, meaning that we observe half of the scientists changing city only once. While, the 12 % of the trajectories are longer (i.e., 455 127). The most frequent trajectories of length one are between

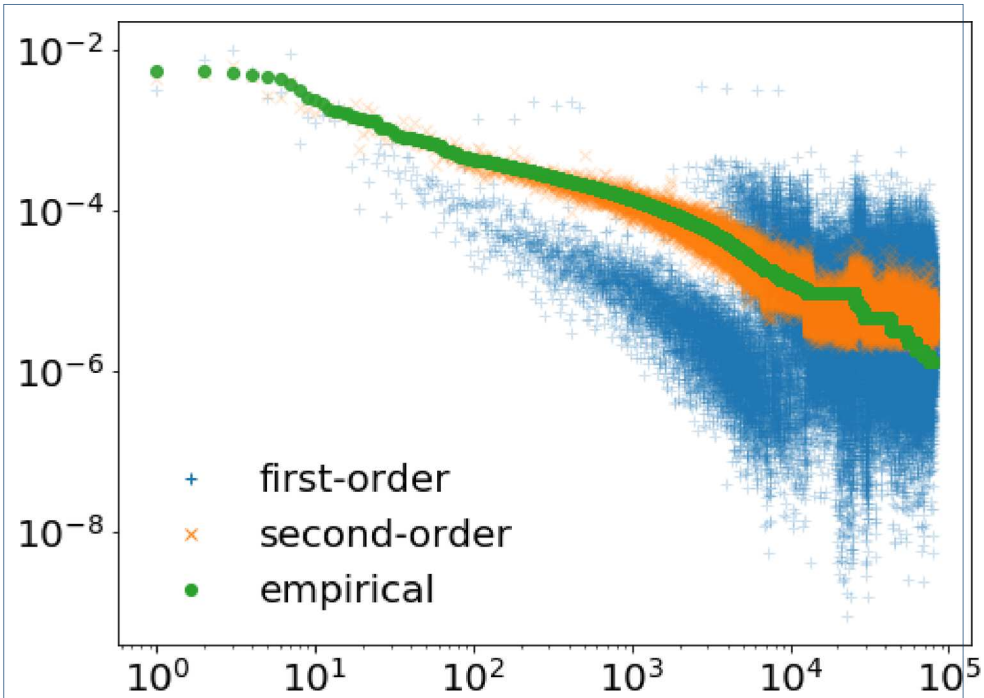

Figure 1 Distribution probabilities for all the motifs of type I and II. we plot $P^{(\mathrm{emp})}$, $P^{(1)}$, and $P^{(2)}$ respectively in green dots, blue $+$ and orange $\times$. The order on the x-axis is created by sorting the motifs from the most to least probable with respect to the empirical data. On the y-axis, we report their respective probabilities.

Boston (MA, USA) and Cambridge (MA, USA), London (UK) and the Oxfordshire (UK), and Tokyo (Japan) and Kanagawa (Japan).

While the most frequent trajectories of length 2 are between Boston (MA, USA), Cambridge (MA, USA) and Boston (MA, USA), Stanford, (CA, USA), Palo Alto, (CA, USA), and Stanford, (CA, USA)[1], London (UK) Oxfordshire (UK), London (UK), Tokyo (Japan), Kanagawa (Japan), Tokyo (Japan).

3 Trajectories at global affiliation level (MAG)

In the MAG, we have over 14 million disambiguated scientist among more than 19000 affiliations. Among these, 2 591 784 trajectories have length 1 or higher between 18 522. Specifically, 1 196 158 are of length 1, meaning that we observe half of the scientists changing affiliation only once. While, 1 395 626 of the trajectories are longer (e.g., 614 776 have length two). The most frequent trajectories of length one are between University of California and University of California, Berkeley University of California and university of California Los Angeles, and University of California and Davis, University of California. Note that University of California is a university system composed of 10 different campuses and is not an unambigu-

[1]Note that the fact that we can distinguish between locations such as Stanford and Paolo Alto only thanks to the fine grain resolution of MapAffil (?). Indeed, by manually checking authors' affiliation that MapAffil placing in Palo Alto, we find that many companies such Xerox, Hewlett-Packard, Lockheed Martin, and many biotech companies have laboratories in Paolo Alto.

ously defined location or affiliation. When considering longer trajectories, we find similar type of trajectories trough ambiguously defined research institutions. For this reasons, we do not use the MAG data to analyze scientists' career trajectories at global level and use MEDLINE instead.

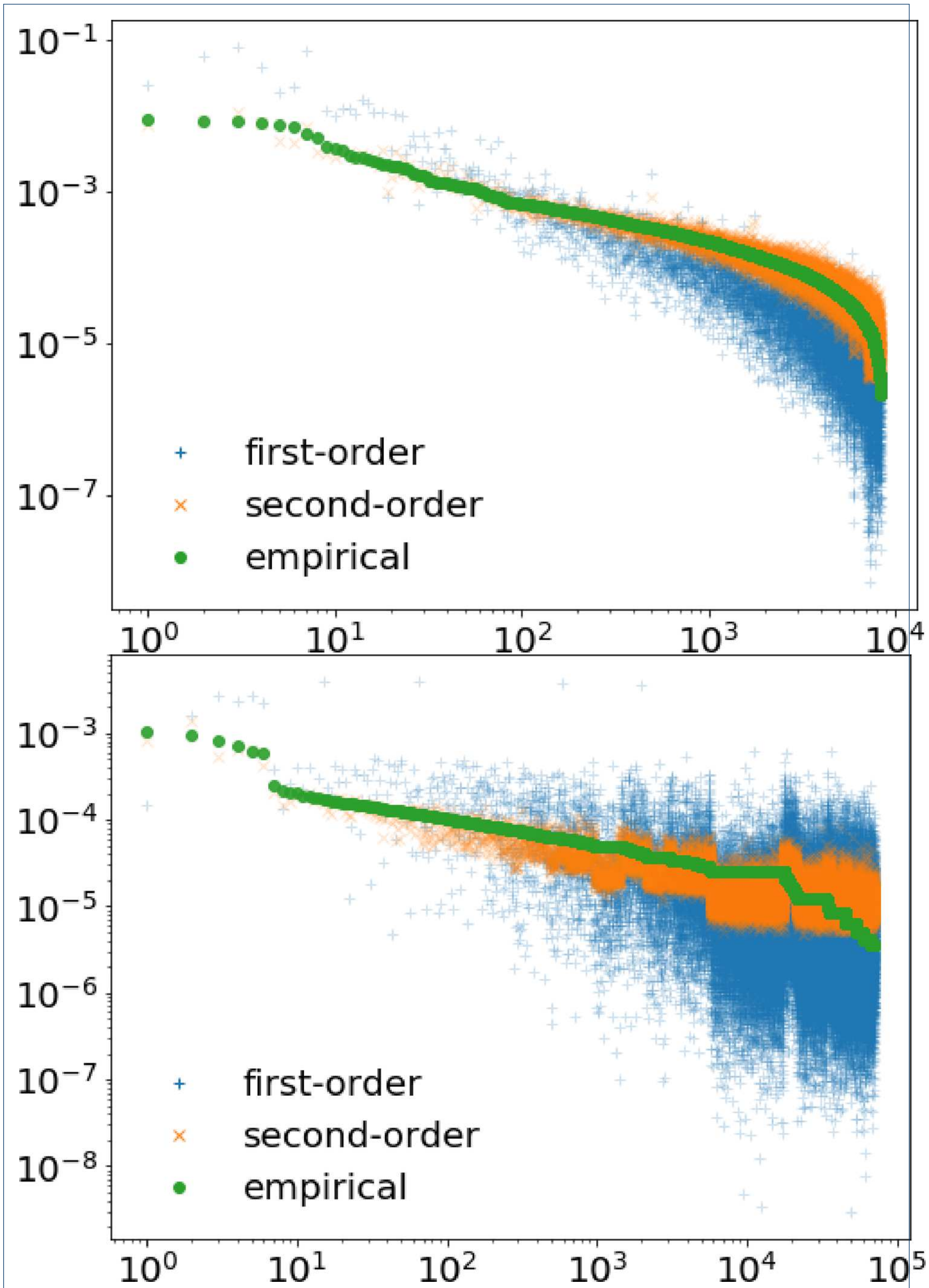

Figure 2 Distribution probabilities for all the motifs of type I (a) and II (b). we plot $P^{(\mathrm{emp})}$, $P^{(1)}$, and $P^{(2)}$ respectively in green dots, blue $+$ and orange $\times$. The order on the x-axis is created by sorting the motifs from the most to least probable with respect to the empirical data. On the y-axis, we report their respective probabilities.