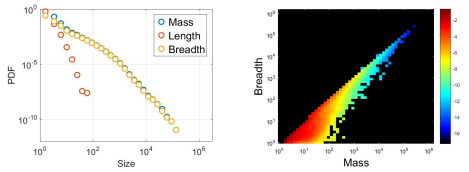
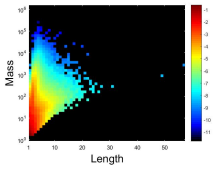
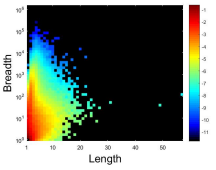


(a) Size

(b) Mass-Breadth

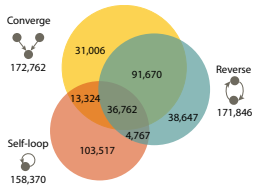


(c) Length-Mass

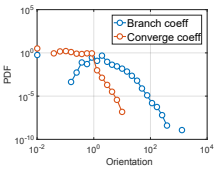


(d) Length-Breadth

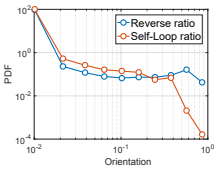
Figure 1: The size metrics for the empirical cascades.



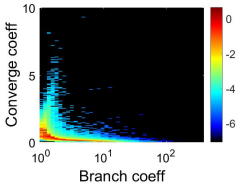
(a) Venn diagram of converge, reverse and self-loop



(b) Branch & Converge



(c) Reverse & Self-Loop



(d) Branch-Converge

Figure 2: The orientation metrics for the empirical cascades.

biggest cascade which is also the widest cascade has mass value $1,414,815$ and breadth value $1,408,024$, and the longest cascade owes the length value 57. However, the average mass, breadth and length values are 5, 4 and 1 respectively.

Size correlation. Figures 1b-d plot the joint density profiles for the size metrics. We find strong positive correlation between mass and breadth as shown in Fig. 1b. Indeed, the correlation coefficient between the logarithmically transformed mass and breadth values is a strikingly high 0.99, indicating that the breadth accounts for a large proportion of mass. Figure 1c plots the joint distribution of length and mass, and we observe the biggest cascades are constrained to a relative small length, while the longest cascades are with moderate mass values. Figure 1d plots the joint distribution of length and breadth. We observe the majority of the cascades are wide and shallow, and there exist narrow and deep cascades. In contrast, it is difficult to find the very wide and deep cascades, or the very narrow and deep cascades.

Orientation. The orientation of a cascade measures to what extent that edges are directionally intertwined within the it. The orientation is characterized by following four metrics: *i*) *Branch coefficient* measures to what extent the edges in C spreading out to different nodes, characterized by the coefficient of variation (the ratio of the standard deviation to the mean) of out-degree distribution $p(k_{out})$ of C, where $k_{out}(u) = \sum_v \mathbb{1}\{(u,v) \in E\}$ is the out-degree of node u and $\mathbb{1}$ is the indicator function. A large branching coefficient value of C means the edges in C spread out from a couple of source nodes to a large amount of destination nodes, implying the orientation of spreading edges are fully random rather than spreading along a preferred direction. *ii*) *Converge coefficient* measures to what extent the edges in cascade C converging into one node, characterized by the coefficient of variance of in-degree distribution $p(k_{in})$ of C, where $k_{in}(v) = \sum_u \mathbb{1}\{(u,v) \in E \ \& \ u \neq v\}$ is the in-degree of node v. A cascade with large converging coefficient value indicates a large proportion of edges pointing to a couple of nodes, implying the information flows tend to converge into few users. *iii*) *Reverse ratio* measures to what extent the edges in cascade C pointing to the reverse direction, characterized by the ratio of the number of reciprocal edges to the total number of edges, i.e., $\frac{|\{(u,v)|(u,v)\in E \ \& \ (v,u)\in E \ \& \ u\neq v\}|}{|E|}$. *iv*) *Self-loop ratio* measures to what extent the edges in C starting and pointing to the same direction, characterized by the ratio of the number of nodes which have self-loop edge to the total number of nodes, i.e., $\frac{|\{u|(u,u)\in E\}|}{|V|}$.

Non-branching orientations. Existing studies of cascade focus mainly on the branching-out orientation, but we find other three orientations are also ubiquitous. Specifically, cascades containing at least one of the converge, reverse or self-loop orientations, shown in Fig. 2a, account for 20.0% of the total population. In addition, cascades usually show a combination of different spreading orientations. The Venn diagram in Fig. 2a shows the number of cascades with different orientation types and their logical relationships. In total, 9.2% of the total cascades have more than two orientation types, as shown in the overlapping region of Fig. 2. The branch coefficient distribution shows a bimodal distribution where two modes are near 0 and 2 respectively, implying that information flow in cascades tend to spread along one direction, or a moderate number of directions. In addition, very large values of branch coefficient do exist. In contrast, the distributions of converge coefficient, reverse ratio and self-loop ratio all peak near 0, a uniform-like distribution at a moderate value range, and followed by a fat-tail range at the large values, implying the prevalence of non-branching orientations and the existence of extreme cases (e.g. each node has a self-loop edge, or each edge has its reverse counterpart).

Orientation correlation. Further, we examine to what extent these spreading orientations can coexist. The branch orientation and converge orientation show non-coexistence relationship like two polarities. Figure 2d plots the heat map of branch coeff. vs. converge coeff. for each cascade. We find that large converge coeff. values only exist with small branch coeff. values, and vice versa.

Acknowledgments Supported by National Program on Key Basic Research Project, No. 2015CB352300; National Natural Science Foundation of China, No. 61370022, No. 61531006, No. 61472444 and No. 61210008, the fund of Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology, National Science Foundation under Grant No. CNS-1314632 IIS-1408924

3. REFERENCES

[1] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *15th ACM SIGKDD*, pages 199–208. ACM, 2009.

[2] D. Liben-Nowell and J. Kleinberg. Tracing information flow on a global scale using internet chain-letter data. *Proceedings of the National Academy of Sciences*, 105(12):4633–4638, 2008.

[3] S.-H. Yang, A. Kolcz, A. Schlaikjer, and P. Gupta. Large-scale high-precision topic modeling on twitter. In *the 20th ACM SIGKDD*, pages 1907–1916. ACM, 2014.

[4] G. Sharad, A. Ashton, H. Jake and W. Duncan The structural virality of online diffusion. In *Management Science*, INFORMS, 2015.

[5] L. Yu, P. Cui, F. Wang, C. Song, and S. Yang. From micro to macro: Uncovering and predicting information cascading process with behavioral dynamics. In *2015 IEEE ICDM*,

[6] C. Zang, P. Cui, and C. Faloutsos. Beyond sigmoids: The nettide model for social network growth, and its applications. In *22nd ACM SIGKDD*, pages 2015–2024. ACM, 2016.

[7] T. Zhang, P. Cui, C. Song, W. Zhu, and S. Yang. A multiscale survival process for modeling human activity patterns. *PloS one*, 11(3):e0151473, 2016.