SocialCache:

A Pervasive Social-Aware Caching Strategy for Self-Operated Content Delivery Networks of Online Social Networks

Tiancheng Guo, Yuke Ma, **Mengying Zhou**, Xin Wang, Jun Wu, Yang Chen

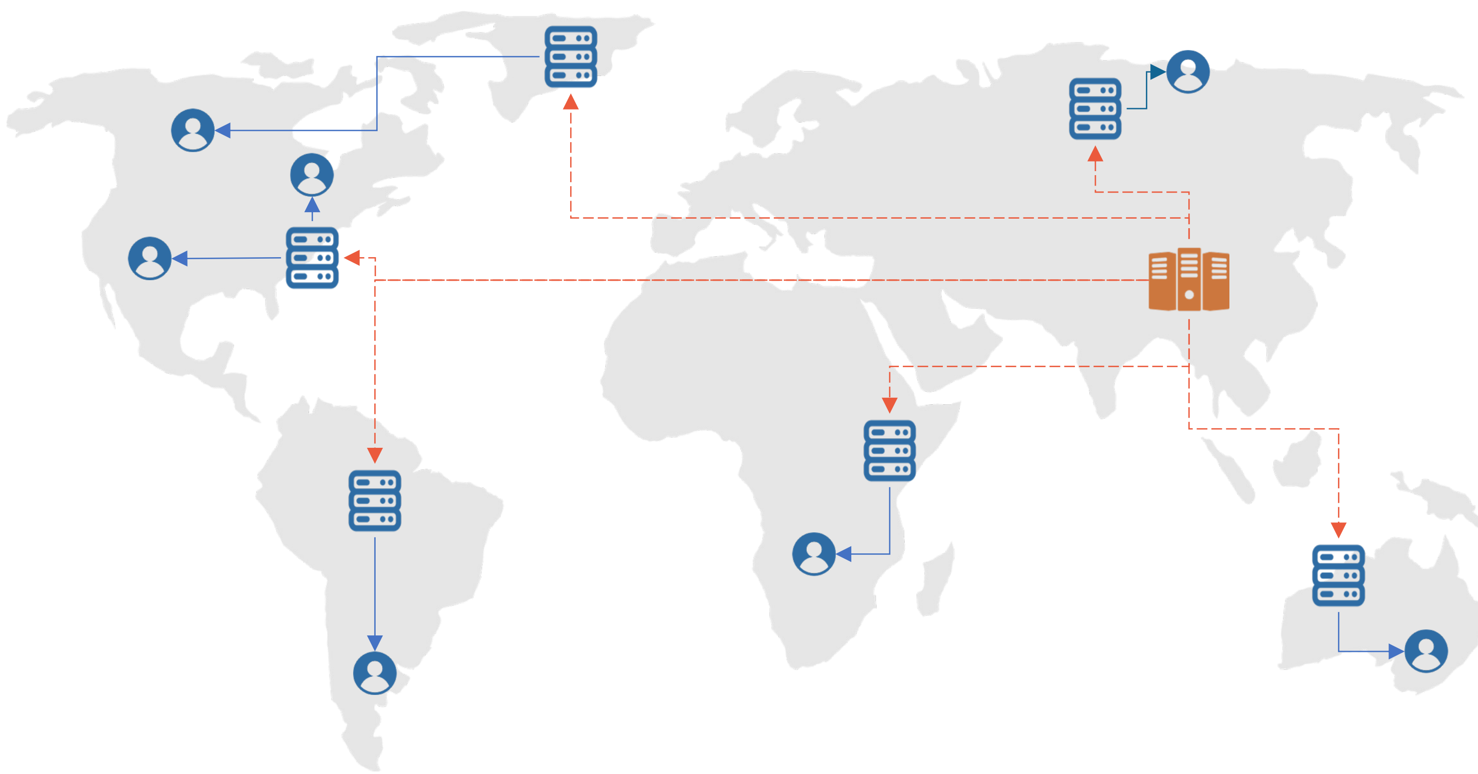
School of Computer Science

Fudan University

Content Delivery Network

CDNs cache content from the origin server on geographically distributed CDN cache servers to reach users faster.

User **CDN server** **Origin server**



Caching Strategy with Traffic Volume

Distributed CDN cache servers are easily to be fully occupied

A

B

C

Full cache

Replace
which content

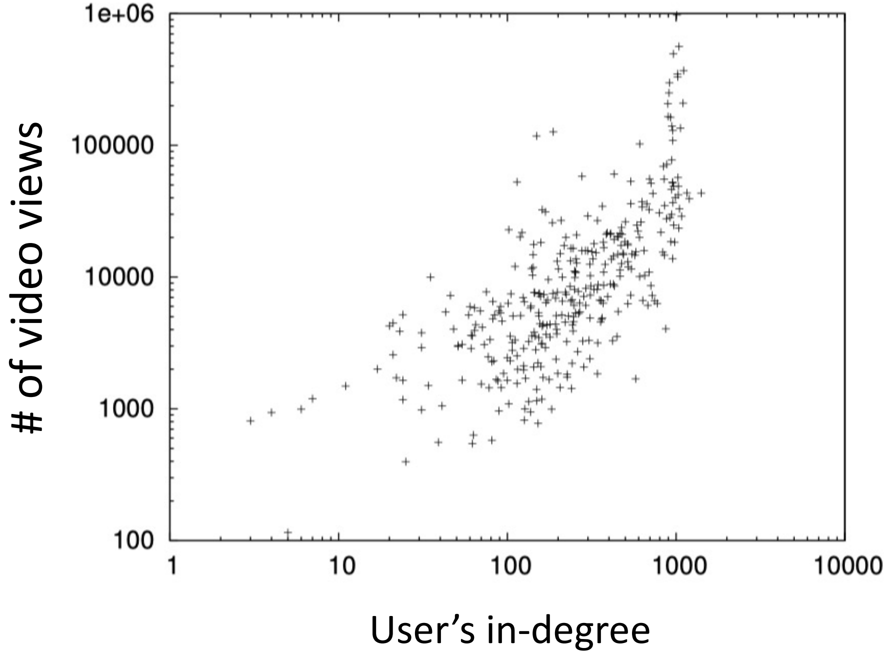
D

New content

Metric: **traffic volume for synchronization**, cache hit ratio, validity period…

Large amount of OSN data

Motivation: Cache Hit and Social Influence

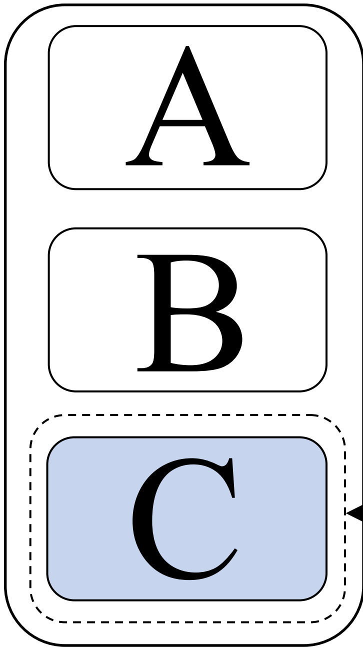


On YouTube, the more followers a user has, the more views their uploaded video content receives [1]

Motivating: Preferentially caching content from users with high social influence can increase the cache hit ratio and reduce traffic for synchronization.

[1] C. Canali, M. Colajanni, and R. Lancellotti. "Characteristics and evolution of content popularity and user relations in social networks." In Proc. of ISCC, 2010.

SocialCache: Social-Aware Caching Strategy



Full cache

Replace cache based on

Priority $\mathcal{P}$

A metric calculated with social connectivity

D

New content

If $min(A,B,C) = C \;\; and \;\; Priority_D > Priority_C$

Then, *D* replace *C* in cache

Calculation of Priority $\mathcal{P}$

$$\text{Calculation of Priority } \mathcal{P} = f\left(\text{Social Connectivity} + \text{Content Size} + \text{Geo Distance} + \text{Timestamp}\right)$$

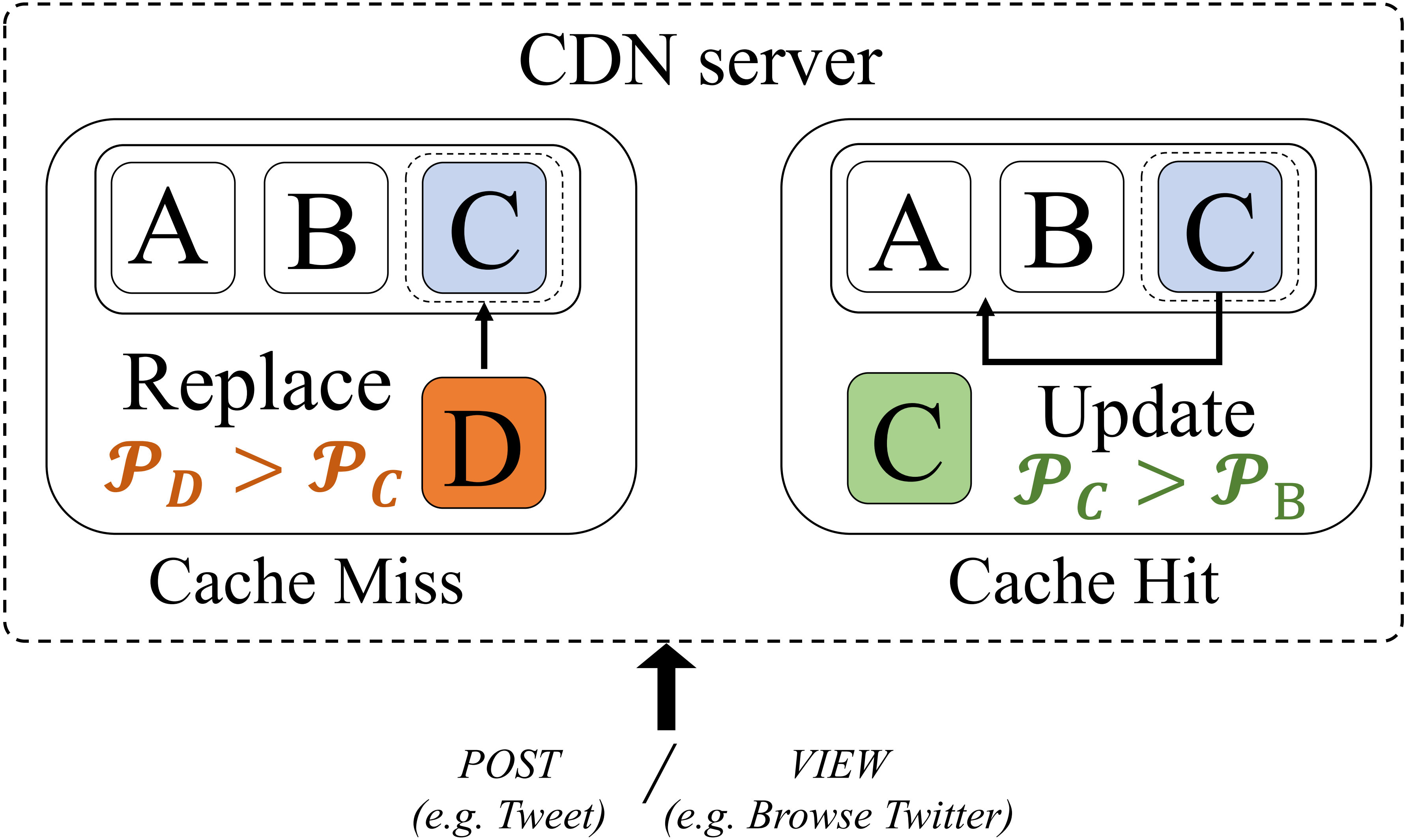
$$= \mathcal{W}_0 \cdot \mathcal{S} + \mathcal{W}_1 \cdot \mathcal{M} + \mathcal{W}_2 \cdot \mathcal{D} + \mathcal{T}$$

- **Social connectivity $\mathcal{S}$**: user's influence in the OSN
- **Content size $\mathcal{M}$**: size of the file transferred by this request
- **Geo Distance $\mathcal{D}$**: geographic distance between the user and the nearest CDN node
- **Timestamp $\mathcal{T}$**: time that this content is created

*We use heuristic algorithm called hill-climbing [1] to fine-tune these weight parameters.

[1] J. Hill and K. Fu, "A learning control system using stochastic approximation for hill-climbing," in Proc. of JACC, 1965.

Two Caching Situations



Evaluation with Real-world Datasets

SocialCache is evaluated on real-world OSN and CDN requests!

Real OSN

- Twitter[1]: 11,088 nodes and 2,420,766 directed edges
- Brightkite[2]: 5,773 nodes and 44,302 edges

Real CDN requests

- CDN requests from Twitter users[3]: 26,952,281 media files

[1] J. J. McAuley and J. Leskovec, “Learning to discover social circles in ego networks,” in Proc. of NIPS, 2012.
[2] E. Cho, S. A. Myers, and J. Leskovec, “Friendship and Mobility: User Movement in Location-Based Social Networks,” in Proc. of KDD, 2011.
[3] Q. Gong, J. Zhang, X. Wang et al., “Identifying Structural Hole Spanners in Online Social Networks Using Machine Learning,” in Proc. of SIGCOMM, Posters and Demos, 2019.

Production and SOTA Baselines

Method	Usage	Temporal info	Social info	Social-aware metric
RAND	Production	×	×	/
FIFO		×	×	/
LRU		√	×	/
LRU-Social	State of the Art (SOTA)	√	√	Susceptible-Infected-Recovered (SIR) spreading model
SocialCache (ours)	***/***	***√***	***√***	***Social connectivity, e.g., effective size, PageRank, Laplacian centrality***

Reduced Network Traffic within CDN

Method	Twitter	Brightkite
RAND	*135.22 GB*	*306.93 GB*
FIFO	124.96 GB	175.15 GB
LRU	124.22 GB	172.41 GB
LRU-Social	124.08 GB	236.55 GB
SocialCache	***116.14 GB*** *(↓ 14.11%)*	***167.28 GB*** *(↓ 45.50%)*

SocialCache can save significant network traffic cost within CDN

*operational costs is $0.085/GB[1]

[1] Z. Zheng, Y. Ma, Y. Liu et al., "XLINK: QoE-Driven Multi-Path QUIC Transport in Large-scale Video Services," in Proc. of SIGCOMM, 2021.

Different Social Connectivity

Social Connectivity Metric	Network Traffic Volume (GB)
In-degree	116.79
PageRank	117.24
Laplacian centrality	117.61
Betweenness centrality	117.52
Effective size	***116.14***

Standard deviation is ***0.54***

Means that different social connectivity has similar performance.

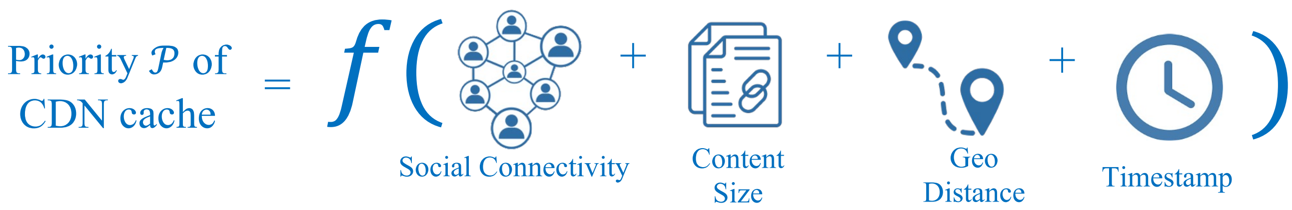
More Considerable Social-Aware

Method	Twitter			Brightkite		
	Network traffic (GB)	Operation Time (s)	Cache Hit Ratio (%)	Network traffic (GB)	Operation Time (s)	Cache Hit Ratio (%)
LRU-Social (SIR model)	124.08	874.65	9.07	236.55	32463.93	48.99
SocialCache (Effective Size)	116.14	48.38 (LRU: 38.71)	10.36	167.28	69.53 (LRU:73.02)	71.81

- **Network traffic and Cache hit ratio:** SocialCache considers geographic location and content size as well, and ignores redundant connections.
- **Operation time:** LRU-social is time-consuming with enumeration of SIR model. SocialCache performs faster and is as efficient as LRU.

SocialCache:

A Pervasive Social-Aware Caching Strategy for Self-Operated Content Delivery Networks of Online Social Networks

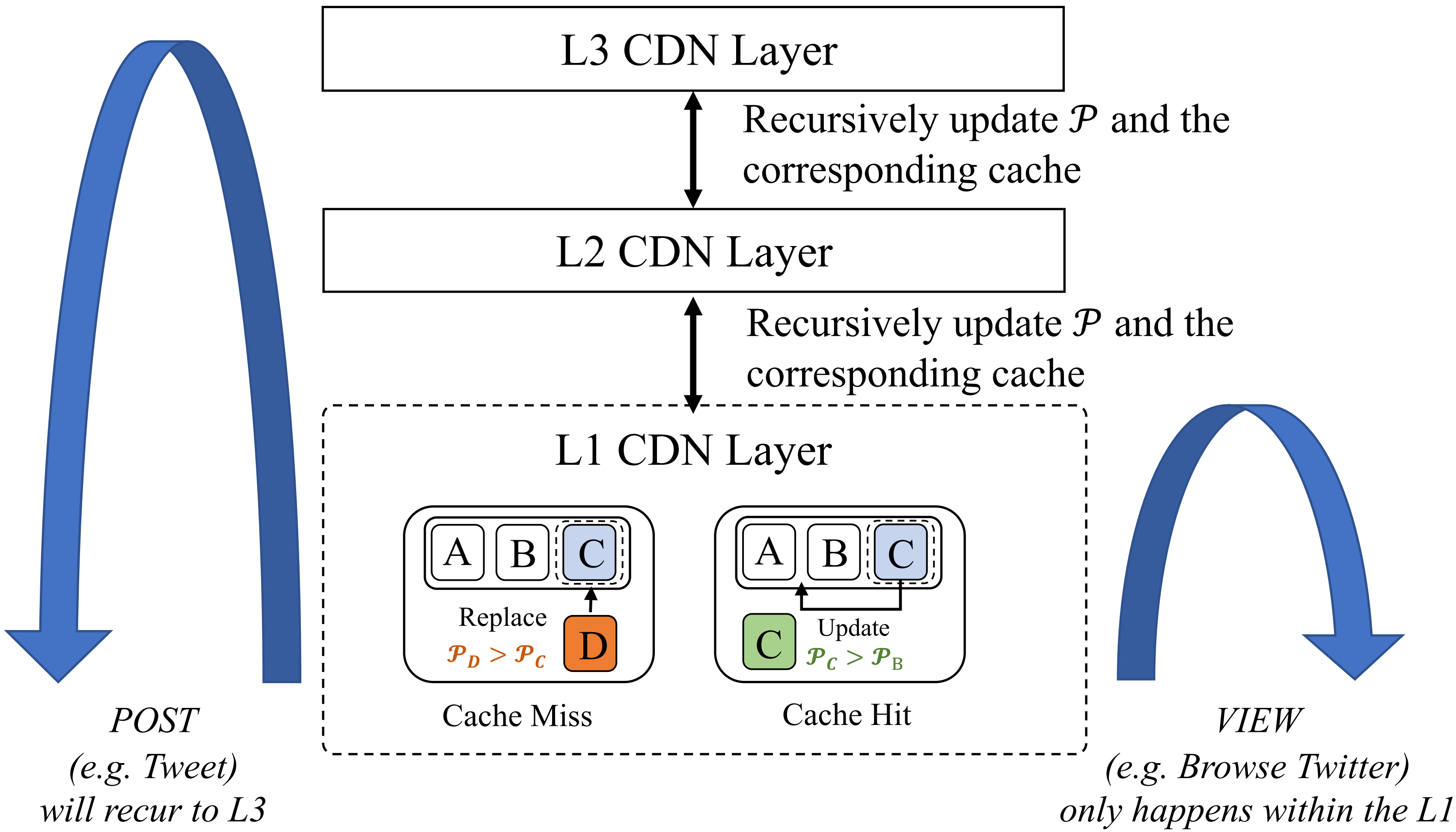


1. We propose SocialCache, a caching strategy to optimize self-operated CDNs with social connectivity information.

2. SocialCache outperforms production and SOTA baselines on real-world OSN and CDN requests datasets, achieving reduced network traffic and operation time.

Thanks for your listening!

Backup Slide – POST and VIEW



Backup Slide: Combine OSN and CDN Requests

Real OSN

- Twitter[1]: 11,088 nodes and 2,420,766 directed edges, 342,542 requests
- Brightkite[2]: 5,773 nodes and 44,302 edges, 749,558 requests.

Real CDN requests

- CDN requests from Twitter users[3]: 26,952,281 media files

For each evaluated dataset

1. # of requests/user: Zipf distribution(α = 1.765, β = 4.888)

$$Zipf(x) = \beta x^{-\alpha}$$

2. Time interval between the requests: LogNormal distribution (μ = 1.789, σ = 2.366)

$$LogNormal(x) = \frac{1}{\sqrt{2\pi}x\sigma} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

[1] J. J. McAuley and J. Leskovec, “Learning to discover social circles in ego networks,” in Proc. of NIPS, 2012.
[2] E. Cho, S. A. Myers, and J. Leskovec, “Friendship and Mobility: User Movement in Location-Based Social Networks,” in Proc. of KDD, 2011.
[3] Q. Gong, J. Zhang, X. Wang et al., “Identifying Structural Hole Spanners in Online Social Networks Using Machine Learning,” in Proc. of SIGCOMM, Posters and Demos, 2019.