# 硕士研究生学位论文开题报告

**论文题目：时空动态的跨网络影响力最大化问题的研究**

**一、**立题依据（包括研究目的、意义、国内外研究现状和发展趋势，需结合科学研究发展趋势来论述科学意义；或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录）（不少于800字）

**1.1 研究背景和意义**

近年来，随着社交网络的兴起，其在信息传播中所扮演的角色也愈加重要。我们发现越来越多的人喜欢在网络上公开他们的观点或想法。用户参与和信息共享的结合使得在线社交网络能够将信息传播到更多的人，并且传播速度甚至比大众媒体更快。

许多学者主要研究兴趣是了解社交网络中的信息传播方式以及如何找到有影响力的用户，例如，一个公司可能希望确定一小部分有影响力的用户，这样他们就可以影响他们的朋友，甚至朋友的朋友，通过口碑效应使得其他用户采用产品。在这样的情况下，公司就会考虑谁可以作为最初的体验用户。这就是所谓的影响力最大化问题，即找到一组用户使得他们在社交网络中的影响范围最广。

然而经典的影响力最大化问题存在着诸多限制。首先它基于以下假设：影响从种子集S传播到社交网络G中的最大节点需要t时间，并且t很小，这样使得G的拓扑在整个传播过程中保持不变。因此，G的拓扑结构在传播过程中是完全已知的。然而在现实网络中，影响传播可能需要相当长的时间(几天甚至几周)才能达到社交网络的最大范围。在这段时间内，网络的拓扑肯定会发生一些变化，比如新用户的加入，原本不认识的用户成为好友等。所以如何使影响力最大化问题的模型和算法能够适应网络的动态变化引起许多研究者的关注。与此同时，随着传播的进行，影响力的大小也应该随着时间的推移发生变化，如何量化影响力大小的衰减过程也是一个值得研究的问题。其次，传统的影响力最大化问题研究的是用户在单一社交网络中的影响力大小。但是在现实中，社交网络往往存在大量的重叠用户，这些用户在跨网络的信息传播方面发挥着重要作用。当用户积极参与多个网络时，他/她可以起到一个转发的作用，同时将信息传播到多个社交网络。如果我们仅仅单独考虑每个网络会大大低估一些用户的影响力。所以如何度量用户在跨网络间的影响力也是一个值得研究的问题。最后，由于社交网络庞大的规模，所以影响力最大化问题的算法应当拥有较低的时间复杂度和空间复杂度才能高效的完成任务。

**1.2 国内外研究现状和发展趋势**

Kempe[1]于2003年首先将影响力最大化建模为一个算法问题。他将一个社交网络抽象为一个图，其中V代表图G中结点的集合（用户集合），E代表图G中边的集合（用户之间的社交关系）。影响力最大化问题的目标是找到一个包含k个结点的用户集合使得它在图G中的影响范围最大。许多商家利用影响力最大化的思想对其商品进行推广，不仅节约了广告成本还使得推广效果更加出众。在影响力最大化问题的研究中，如何对信息扩散进行建模和怎样选择种子用户是现在主要的研究点。

**1）信息扩散过程模型**

对信息扩散过程进行建模是评估种子集合影响范围的重要基础[2]。最经典的两类传播模型是独立级联模型（ICM）和线性阈值模型（LTM）。

独立级联模型最早由J. Goldenberg[3]于2001年中提出。传统的独立级联模型的主要思想是当一个节点u被激活时，它会以概率p(u,v)对它未激活的出边邻居节点v尝试激活，这种尝试仅仅进行一次，而且这些尝试之间是互相独立的，即u对v的激活不会受到其他节点的影响。它的传播过程如下：

·给定初始的活跃节点集合S，当在时刻t节点u被激活后，它就获得了一次对它的邻居节点v产生影响的机会，成功的概率为p(u,v)，是随机赋予的系统参数。

·若v有多个邻居节点都是新近被激活的节点，那么这些节点将以任意顺序尝试激活节点v。如果节点u成功激活节点v，那么在t+1时刻，节点v转为活跃状态。

·在t+1时刻，节点v将对其他节点产生影响，重复上述过程。

·结束条件：网络中不存在有影响力的网络节点。

线性阈值模型最早由Granovetter 和Schelling[4]于1978年提出。传统线性阈值模型的主要思想是对图中每个节点v都分配一个激活阈值，当v的入邻居影响力大于它的阈值时，就会被激活。

传统线性阈值模型的传播过程如下：

·集合中的任意节点v随机分配阈值θ[v]∈[0,1]。只有当节点v的新处于激活状态的邻居节点对它的影响力大于该阈值时，节点v才能被激活。

·用权值b[u.v]表示节点v被它的邻居节点u的影响，∑u∈in(v)  b[u,v]≤1表示节点v的处于活跃状态的邻居节点对它的影响力之和。这里in(v)是v的入边邻居节点集合。

·给定初始的活跃节点集合S，在t时刻，所有在t-1时刻处于活跃状态的节点仍保持活跃，并且当这一时刻节点v的邻居节点的影响力之和大于节点b的阈值时，节点v被激活.

·节点v被激活后，下一时刻将对它的邻居节点产生影响，重复上述过程。

·结束条件：当网络中已存在的所有活跃节点中任意活跃节点的影响力之和都不能激活他们的处于非活跃状态的邻居节点时。

**上述两个基本模型虽然可以大致描述信息在社交网络中的传播过程，但是存在很多的局限性。**

**首先，随着在线社交网络的发展，越来越多的用户会同时使用多个社交网站，如果只在单一网站上衡量其影响是不妥的**

**对传统传播模型的改进还体现在将其扩展到多个网络之中**。Nguyen D T等人[10] 首先提出了如何评估影响同时在多个网络中的传播。他提出一种将多个网络耦合成一个新的单一网络的方案，该方案保留了每个网络的影响扩散特性。随后，Nguyen D T等人[11]又提出了在多个网络中的最小成本影响问题（LCI），该问题要求确定最终能够影响大量用户的种子用户的最小数量，作者解决了多路OSNs中的LCI问题，通过不同的耦合方案将多路网络简化为一个网络，同时保留了影响最大的传播特性。Zhan Q[12]认为在现实世界中，在线社交网络通常是异构的，用户可以通过多个渠道相互影响，并且在线社交网络还可以共享公共用户，信息可以通过这些用户在不同网络中传播。本文首次研究了部分对齐异构社交网络中的影响最大化问题，并且提出了MM模型，它基于一组网络内部和网络内部的社会元路径，从对齐的异构社交网络中提取多对齐的多关系网络(MMNs)。Zhan Q [13]认为如今的用户通常同时参与多个在线社交网络，那些加入Facebook的用户也在使用其他网络，比如Twitter，Foursquare和Instagram。如果想要研究Facebook网络中影响力问题的话，也要研究影响力从上述其他网络中的传入。Ma Y[14] 认为本地专家可以在很多应用中发挥重要作用，如解决本地信息查询、社会事件安排等。作者在多个社交网络中寻找本地专家，并且用到了两个指标：local authority和topic authority。这两个指标分别与地理位置和主题相关。Kuhnle A[15]认为影响在不用的网络中的传播模型是不一样的。作者将不同网络抽象为一层，并且提出了在不同层中传播模型可能不一样的传播方式。Singh[16]认为在多个网络中，种子用户可以同时宣传多个产品，而非种子用户可以同时在网络上接受不同的产品。针对上述场景，作者提出了一个跨多个社交网络的多重影响最大化框架(MIM2)。

**其次，随着科技的发展，位置社交网络逐渐兴起，用户的位置对于信息的传播也有着很大的影响。**

**Li, G**将带地理位置的影响力最大化定义为：给定一个查询Q = (R, k)，一个地理区域R，一个整数k，令VR表示R中的顶点集合，作者的目标是在图中找到k个节点组成的集合S使得其可以影响到VR中最多的节点。**Wang, Xiaoyang**提出了距离感知的影响力最大化问题，并且给出了距离感知的影响传播如下：给一个地理社交网络G=(V,E)和一个地理查询q，节点集合S的距离感知的影响传播为Iq(S),被计算为其中w(v，q)是v相对于q的权重，I(S,v)是集合S激活v的概率。**Wang, L**将地理位置与主题信息相结合，认为用户之间的传播概率不仅与他们的空间接近度相关，还与两者之间的话题兴趣度相关。此外作者还考虑了选择种子用户时的成本，用PageRank来表示种子用户的激活成本。康瑜师姐在《InDes-TPLP: Two-Phase Location Promotion Algorithm Combining with User Interests and Spread Desires》中考虑了签到移动性，位置自身吸引力，用户签到相似性以及传播者影响力来度量用户之间的传播概率。

**最后，随着传播的进行，影响的强度也是会逐渐下降的。**

**Liu等人**[5]认为一个用户影响另一个用户所需要的时间是不同的并且提出了时间延迟感知的独立级联模型（LAIC），在这个模型下，每个结点的状态分为三种，激活，延迟激活，非激活。在LAIC模型中，当一个节点u在t步第一次被激活时，他会在步骤对它的当前不活跃节点v以概率进行激活，其中是一个影响延迟。**Kim等人**[6]认为尝试激活的次数不应该只有一次所以提出了连续激活和时间限制的IC模型（CT-IC），该模型的传播方式为：给定一个有向图G=（V，E），一个种子集合S，和一个时间限制T，使用At表示在时间t的活跃种子，在时间t，所有属于At的节点都要以概率试图去激活它的未活跃邻居。**Chen 等人**[8]认为影响从一个人到另一个人的传播是需要二者之间“相遇”才有机会传播的。这里所谓的相遇不一定非要是两个人面对面在一起，也可以是在社交媒体上的点赞转发，于是他提出了IC-M模型。**Li H等人**[9] 认为经典的影响力最大化问题忽略了影响传播的时间，而影响传播的时间在现实中可能是非常重要的，在此期间，潜在的社交网络可能会进化。作者假设传播符合独立级联模型（ICM），并且结合了Forest Fire model（森林火灾模型）来预测从G0到Gr的网络拓扑结构。**Ohsaka等人**[28]提出了Time-Varying IC Model。在TVIC模型中，假设一开始有一个种子集合A，随后按照以下规则进行传播：当节点u在时间tu被激活后，他有一次机会通过边e(u,v)向他的非激活邻居v进行激活。这里不同于标准IC模型，到v的距离和激活v的概率都依赖于时间。即影响在t时刻达到v的条件概率由fe(t | tu)定义。这里有fe(t | tu)=fe(t-tu)。此外，当v在t时刻受到影响时，被激活的概率由一个非递增函数pe给出。因此，v在t时刻被激活的概率是

Pr[v becomes active at time t|u is active at time tu]=pe(t)fe(t-tu)。**Liqing[29]**等人对上述liu提出的LAIC模型进行了改进，将liu提出的泊松分布和几何分布换成了幂律分布。

**2）对动态网络结构进行预测**

作者在下列文献中提出（Goyal, P., Chhetri, S. R., & Canedo, A. (2019). dyngraph2vec: Capturing network dynamics using dynamic graph representation learning. Knowledge-Based Systems.）提出了dyngraph2vec预测模型，其目标是捕获进化的潜在网络动态。给定图的时间快照，我们的目标是在每个时间步上学习节点的表示，同时捕获动态，以便预测它们未来的连接。dyngraph2vec使用多个非线性层来学习每个网络中的结构模式，此外，它使用递归层来学习网络中的时间转换。重复层中的回看参数控制学习时间模式的长度

作者在下列文献中提出（Goyal P, Kamra N, He X, et al. Dyngem: Deep embedding method for dynamic graphs[J]. arXiv preprint arXiv:1805.11273, 2018.）了一种高效的动态图嵌入算法DynGEM，用于生成动态图的稳定嵌入。DynGEM在其核心使用了一个深度自动编码器，并利用深度学习的最新进展来生成高度非线性的嵌入。DynGEM不是从头开始学习每个快照的嵌入，而是从t - 1时刻的快照逐步构建t时刻的快照的嵌入，具体地，我们从之前的时间步长初始化嵌入，然后进行梯度训练。这种方法不仅保证了嵌入的稳定性，而且由于在第一个时间步之后的所有嵌入都需要很少的迭代来收敛，因此也带来了有效的培训。

作者在下列文献中提出（Taheri, A., Gimpel, K., & Berger-Wolf, T. (2019, May). Learning to Represent the Evolution of Dynamic Graphs with Recurrent Models. In Companion Proceedings of The 2019 World Wide Web Conference (pp. 301-307). ACM.）了一种动态图的无监督表示学习架构，旨在学习随时间演化的图的拓扑和时间特征。该方法由一个嵌入门控图神经网络(GGNNs)和长短时记忆网络(LSTMs)的序列-序列编译码器模型组成。GGNN能够在每个时间步上学习图的拓扑结构，而LSTMs则利用时间步来传播时间信息。此外，编码器学习演化图的时间动态，而解码器利用编码器提供的编码表示，在同一时间段内重建动态。

**3）种子用户选择算法**

尽管影响力最大化问题在计算上很复杂并且已经被证明是个NP-hard问题[1]，但是依然可以通过一些巧妙的算法来得到近似最优解。得到最优解的条件是影响函数要满足以下两个条件：单调性和子模性。直观上来说单调性就是在种子集合S中加入一个结点后，它的影响范围不会减少，而子模性的意思是在集合S中加入一个结点时，这个结点所带来的边际效益要小于上一个结点的加入带来的边际效益。D. Kempe等人[1]已经证明了影响函数在经典的IC模型和LT模型中是满足上述两个性质的。大量社交网络的影响力最大化算法是基于贪心算法和启发式算法。D. Kempe等人[1]提出了使用蒙特卡洛模拟来评估每个结点的边际收益，并且每次将边际收益最大的结点加入到种子集合中。这样的做法在网络比较大的时候显然是非常低效的。为了解决这个问题，J. Leskovec等人[17]提出了CELF算法来减少蒙特卡洛模拟的次数，它首先计算所有的结点的边际收益，得到初始集合S1。前一次得到的边际收益就是下一次的上界。它对余下结点按照上一次计算的边际收益的降序来进行计算，如果一个节点的边际收益值近似于上界，则将其直接加入集合中。C. Zhou等人[18]提出了UBLF，用了一种矩阵分析的方法快速获取所有结点的边际收益上界。Wang等人提出了CGA算法，它将图划分成社区后利用社区内每个节点的影响来决定选择哪些节点作为种子，它只用在局部子图中运行蒙特卡洛从而降低蒙特卡洛模拟的复杂度。上述使用蒙特卡洛模拟的方法虽然可有很好的模型通用性，但是却阻碍了性能的提升。因此，最近研究人员已经开始探索基于代理和基于草图的方法。影响力排名代理的想法非常直观。它根据一个近似用户影响的度量对图G中的所有用户进行排序，然后直接从排序中生成种子集。因此，这里最重要的挑战是得到一个好的排名指标[2]。比较经典的一个算法是W. Chen等人[19]提出的MIA算法，它的主要思想是计算节点u对植根于u的局部树状结构的影响扩散。基于草图的方法在提高理论效率的同时也可以使近似值得到保证。基于草图的方法基于特定的扩散模型预先计算多个草图，然后利用草图来评估影响扩散。它一般分为两大类，正向影响草图（FI-Sketch）和反向可达草图（RR-Sketch）。正向影响草图与蒙特卡洛方法相似。在一个图中，假设每条边上的传播概率为P1,P2…Pn。利用蒙特卡洛模拟计算一个节点的影响范围时，首先将图中所有边以1-Pk的概率移除，在这个子图中计算当前节点的影响范围。S. Cheng等人提出的staticGreedy算法建立个草图并计算影响范围。但是正向影响草图在最坏情况下的时间复杂度依然很高。反向可达草图通过选择随机结点并生成随机的RR集来估计结点的边际收益。Y. Tang等人[21]提出的TIM算法可以确定到底需要多少RR集但是对内存的消耗很大。为了减少内存浪费X. Wang等人[22]提出了BKRIS算法，它首先用启发式的方法去估计一个OPT的下限，然后从这个下限得出数量足够多的RR集。

然而上述算法都没有考虑到网络的动态性。在现实世界中，社交网络具有高度的动态性，并且随着时间的推移发展迅速[23]，因此，关于影响力结点的计算和结果很快就会过时。Zhuang[24]考虑了一个社交网络在动态的变化，但是只能在特殊时间探测到少数结点，因此，信息只在当前网络的一小部分可用。文献[25][26][27]都使用一系列网络快照来表示图在一定时间内的动态变化。

目前，几乎没有人在跨网络传播中考虑时间因素，跨网络传播中的动态问题更是几乎没有提及。由此可见，对影响在跨网络中的动态传播进行建模并且提出一种算法找到最具影响力的用户是很有必要的。

参考文献

1. D. Kempe, J. Kleinberg, and E. Tardos, “Maximizing the spread of influence through a social network,” in Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2003, pp. 137–146
2. Li, Y., Fan, J., Wang, Y., & Tan, K. L. (2018). Influence maximization on social graphs: A survey. IEEE Transactions on Knowledge and Data Engineering, 30(10), 1852-1872.
3. J. Goldenberg, B. Libai, and E. Muller, “Talk of the network: A complex systems look at the underlying process of word-ofmouth,” Marketing Lett., vol. 12, no. 3, pp. 211–223, 2001
4. M. Granovetter, “Threshold models of collective behavior,” Am.J. Sociol., vol. 83, no. 6, pp. 1420–1443, 1978.
5. Liu, Bo, et al. "Time constrained influence maximization in social networks." 2012 IEEE 12th international conference on data mining. IEEE, 2012
6. Kim, Jinha, Wonyeol Lee, and Hwanjo Yu. "CT-IC: Continuously activated and time-restricted independent cascade model for viral marketing." Knowledge-Based Systems 62 (2014): 57-68.
7. Mohammadi A, Saraee M, Mirzaei A. Time-sensitive influence maximization in social networks[J]. Journal of Information Science, 2015, 41(6): 765-778.
8. Chen, Wei, Wei Lu, and Ning Zhang. "Time-critical influence maximization in social networks with time-delayed diffusion process." Twenty-Sixth AAAI Conference on Artificial Intelligence. 2012.
9. Li H, Bhowmick S S, Cui J, et al. Time is what prevents everything from happening at once: Propagation time-conscious influence maximization[J]. arXiv preprint arXiv:1705.10977, 2017.
10. Nguyen D T, Das S, Thai M T. Influence maximization in multiple online social networks[C]//2013 IEEE Global Communications Conference (GLOBECOM). IEEE, 2013: 3060-3065.
11. Nguyen, D. T., Zhang, H., Das, S., Thai, M. T., & Dinh, T. N. (2013, December). Least cost influence in multiplex social networks: Model representation and analysis. In 2013 IEEE 13th International Conference on Data Mining (pp. 567-576). IEEE.
12. Zhan Q, Zhang J, Wang S, et al. Influence maximization across partially aligned heterogenous social networks[C]//Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Cham, 2015: 58-69.
13. Zhan Q, Zhang J, Philip S Y, et al. Discover tipping users for cross network influencing[C]//2016 IEEE 17th International Conference on Information Reuse and Integration (IRI). IEEE, 2016: 67-76.
14. Ma Y, Yuan Y, Wang G, et al. Local Experts Finding Across Multiple Social Networks[C]//International Conference on Database Systems for Advanced Applications. Springer, Cham, 2019: 536-554.
15. Kuhnle A, Alim M A, Li X, et al. Multiplex influence maximization in online social networks with heterogeneous diffusion models[J]. IEEE Transactions on Computational Social Systems, 2018, 5(2): 418-429.
16. Singh, S. S., Singh, K., Kumar, A., & Biswas, B. (2019). MIM2: Multiple influence maximization across multiple social networks. Physica A: Statistical Mechanics and its Applications, 526, 120902.
17. J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, “Cost-effective outbreak detection in networks,”in Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2007, pp. 420–429
18. C. Zhou, P. Zhang, W. Zang, and L. Guo, “On the upper bounds of spread for greedy algorithms in social network influence maximization,” IEEE Trans. Knowl. Data Eng., vol. 27, no. 10, pp. 2770–2783, Oct. 2015.
19. W. Chen, C. Wang, and Y. Wang, “Scalable influence maximization for prevalent viral marketing in large-scale social networks,” in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 1029–1038
20. S. Cheng, H. Shen, J. Huang, G. Zhang, and X. Cheng,“Staticgreedy: Solving the scalability-accuracy dilemma in influence maximization,” in Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage., 2013, pp. 509–518.
21. Y. Tang, X. Xiao, and Y. Shi, “Influence maximization: Nearoptimal time complexity meets practical efficiency,” in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2014, pp. 75–86.
22. X. Wang, Y. Zhang, W. Zhang, X. Lin, and C. Chen, “Bring order into the samples: A novel scalable method for influence maximization,” IEEE Trans. Knowl. Data Eng., vol. 29, no. 2, pp. 243–256, Feb. 2017.
23. R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In Link Mining: Models, Algorithms, and Applications, pages 337-357.Springer, 2010
24. Zhuang, H., Sun, Y., Tang, J., Zhang, J., & Sun, X. (2013, December). Influence maximization in dynamic social networks. In 2013 IEEE 13th International Conference on Data Mining (pp. 1313-1318). IEEE.
25. X. Chen, G. Song, X. He, and K. Xie, “On influential nodes tracking in dynamic social networks,” in Proc. SIAM Int Conf Data Mining, 2015, pp. 613–621.
26. Xu, S., Xu, N., Zhang, J., Li, F., & Li, S. (2017, December). Seed set selection in evolving social networks. In 2017 3rd IEEE International Conference on Computer and Communications (ICCC) (pp. 2323-2328). IEEE.
27. Hafiene N, Karoui W, Romdhane L B. Influential Nodes Detection in Dynamic Social Networks[C]//International Conference on Business Information Systems. Springer, Cham, 2019: 62-73.
28. Ohsaka, N., Yamaguchi, Y., Kakimura, N., & Kawarabayashi, K. I. (2016, September). Maximizing time-decaying influence in social networks. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 132-147). Springer, Cham.
29. Liqing, Q., Jinfeng, Y., Xin, F., Wei, J., & Wenwen, G. (2019). Analysis of Influence Maximization in Temporal Social Networks. IEEE Access, 7, 42052-42062.

**二、研究内容和目标（说明课题的具体研究内容，研究目标和效果，以及拟解决的关键科学问题。此部分为重点阐述内容）（不少于2500字）**

**2.1 研究内容**

在单一位置社交网络无法很好的反映用户真实影响力的情况下，通过构建跨位置社交网络的影响扩散模型，利用其他成熟社交网络的信息作为辅助，从而准确反映用户的影响力。同时，为了满足随着时间推移，网络拓扑结构的变化以及用户地理位置的变化，我们对时空动态的社交网络的信息扩散过程进行建模，通过预测某一时刻网络的拓扑结构和用户地理位置来解决目前对网络的动态性预测不准的问题。最后针对现有的种子选择算法进行改进，提出一种高效的种子选择方法。从而为商家提供更准确，更高效的位置推广服务。

**（1）对跨位置社交网络的信息扩散过程进行建模**

**1）在位置社交网络（目标网络）中如何对用户之间的传播概率进行度量**

网络中用户间的信息扩散概率反映了待推广商铺的信息将有多大的概率从源用户传播到目标用户。在位置社交网络中，对于一个商家进行推广时，信息的扩散概率与用户的历史访问兴趣，被推荐商铺对用户的吸引力以及用户之间的签到相似度都有关系。如何利用上述信息并且将其进行聚合从而精确度量出位置社交网络中的两个用户之间的信息扩散概率是我们课题研究内容的难点之一。

**2）在社交网络（源网络）中如何对用户之间的传播概率进行度量。**

在跨网络的传播中，我们使用一个社交网络的辅助信息从而可以更好的度量用户的实际影响力。信息在源网络中的传播与在目标网络中的传播是互相独立进行的。由于普通的社交网络中没有关于用户的更多信息。如何度量信息在源网络中的传播才能更好的反映用户的影响力也是一个值得思考的问题。

**3）对信息在两个网络中的扩散过程进行描述。**

现有的关于信息扩散模型的研究主要是基于两大经典模型：独立级联模型和线性阈值模型。在独立级联模型中用户之间的激活是独立的，而在线性阈值模型中用户是否激活是受到他所有邻居的的共同影响，可以看成是一种价值累积模型。

不同于单个网络，信息在多个网络中的传播是一件很复杂的事情。信息不光可以单独在各个网络中扩散，还可以通过锚用户和锚链接在多个网络中来回扩散。所以如何对信息在跨网络中的传播进行建模也是我们的研究内容之一。

**（2）对时空动态的网络进行预测**

**1）通过现有的网络结构对未来某时刻的网络结构进行分析和预测。**

随着时间的推移，多源位置社交网络的结构也是在动态变化的，网络中可能有节点的加入和退出，也会有节点间关系的建立与删除，也会有新增的签到等信息。此外，目前城市的人口流动性很强，导致商家需要考虑谁在未来某一时刻的影响力最大。本课题拟根据目前网络的结构对未来某一时刻的网络进行预测，并且以此来表示整个网络在一段时间内的动态变化。使用何种方法可以使得预测结果最接近真实网络是我们研究的难点之一。

**（3）如何度量用户的影响力并选出影响力最大的一组用户**

**1）针对待推广的商家，如何在其周围选定一块区域，使得这片区域中被激活的用户去商家签到的可能性最大。**

在对一个商家进行推广的时候，并不是用户对这个商家产生兴趣就一定会去，用户到商家的距离也是要考虑的一大问题。比如我们在看一个美食推送的时候，会觉得这个看上去很好吃，可是当我们想去的时候，一看离我们住的地方很远，立马就会萌生退意。所以如何在被推广商家周边划分一个区域，使得这个区域内的用户一旦表现出对商家的兴趣就会有很大的概率去“打卡”是我们要研究的一个重要问题。

**2）判断用户是否还在查询区域内**

随着时间的推移，用户的位置也有可能发生变化。比如在初始网络中，用户处于查询区域内，但是经过一段时间之后，用户已经不在查询区域内了，这时候我们要计算某个用户在一个查询区域内的影响力时就不能将其计算在内了。所以如何判断一个用户是否在查询区域内是我们要研究的内容之一。

**3）找到合理的种子选择算法**

目前有两种主流的种子选择思想。一种是基于启发式的思想，这种方法使用预先定义的选种策略进行选种，复杂度低，选种速度很快，但是它有一个致命的缺点就是扩散规模很小所以无法准确的度量一个用户影响力的大小。另一种是基于贪心的思想，这种方法每次选择当前影响力最大的种子加入到种子集合中，时间复杂度较高，但是扩散效果很好，可以很好的度量用户影响力的大小。

如何设计一个种子选择算法使得在异构社交网络中具有最大的扩散规模同时又可以降低算法的时间复杂度是本课题面临的一个重要挑战。目前最性能最好的影响力最大化算法是反向可达草图（RR）的方法。我将基于它进行改进。

除此以外，随着传播的进行，网络的拓扑结构也在变化，如何动态的将种子集中的种子用户进行更新也是研究的内容之一。

**2.2研究目标**

本课题的研究目标是在单一位置社交网络无法真实得反映用户影响力并且现实中人口流动过大的情况下，通过将位置社交网络与其他社交网络进行融合来找到新的信息扩散途径，并且对多源网络进行预测，并基于此构建跨位置社交网络的动态影响扩散模型以及跨网络的高效种子用户选择算法

（1）基于给定的位置信息和构建好的多源位置社交网络，构建跨位置社交网络的信息扩散模型。该阶段的具体研究目标为：

1）在位置社交网络（目标网络）中如何对用户之间的传播概率进行度量

2）在社交网络（源网络）中如何对用户之间的传播概率进行度量。

3）对信息在两个网络中的扩散过程进行刻画。

（2）基于给定的位置信息和构建好的多源位置社交网络，构建时空动态的社交网络的信息扩散模型。本阶段的具体研究目标为:

1）为了得到未来某个时刻的网络拓扑结构，需要对现有的网络结构进行分析和预测。使用何种方法可以使得预测结果最接近真实网络是我们研究的难点之一。

（3）基于上述动态跨网络的信息扩散模型度量用户的影响力并提出一种高效的种子选择算法

1）针对待推广的商家，如何在其周围选定一块区域，使得这片区域中被激活的用户去商家签到的可能性最大。

2）如何判断用户是否还在查询区域内

3）找到合理的种子选择算法

如何设计一个种子选择算法使得在异构社交网络中具有最大的扩散规模同时又可以降低算法的时间复杂度是本课题面临的一个重要挑战。目前最性能最好的影响力最大化算法是反向可达草图（RR）的方法。我将基于它进行改进。

除此以外，随着传播的进行，网络的拓扑结构也在变化，如何动态的将种子集中的种子用户进行更新也是研究的内容之一。

**2.3拟解决的关键科学问题**

**三、研究方案设计及可行性分析（包括：研究方法，技术路线，理论分析、计算、实验方法和步骤及其可行性等）（不少于800字）**

**四、本研究课题可能的创新之处（不少于500字）**

**五、研究基础与工作条件（1.与本项目相关的研究工作积累基础 2.包括已具备的实验条件，尚缺少的实验条件和拟解决途径）（不少于500字）**