



单位代码 10006

学 号 10061016

分 类 号 TP301.6

北京航空航天大学

B E I H A N G U N I V E R S I T Y

毕业设计 (论文)

小世界网络中复杂传播和复杂路由的研究

学 院 名 称 计算机学院

专 业 名 称 计算机科学与技术

学 生 姓 名 李强

指 导 教 师 任爱华 陈卫

2014 年 06 月

小世界网络中复杂传播和复杂路由的研究

李强

北京航空航天大学

北京航空航天大学

本科生毕业设计（论文）任务书

I、毕业设计（论文）题目：

小世界网络中复杂传播和复杂路由的研究

II、毕业设计（论文）使用的原始资料（数据）及设计技术要求：

本文使用的原始资料（数据）：在微软亚洲研究院实习时已有的一些实验数据和每次的会议记录

本文的设计技术要求：熟练阅读外文文献，了解小世界网络的理论知识，在证明过程中用到了概率论、图论和计算复杂性相关知识。

III、毕业设计（论文）工作内容：

首先，进行蒙特卡洛模拟，进行 Kleinberg 小世界网络上的模拟实验。

其次，给出复杂传染病的传播理论上的分析，提升已有的复杂传播时间的下界。

然后，研究一种新的消息扩散行为，复杂传染病的路由，并分析复杂路由的效率。

最后，分析对比简单传播、简单路由、复杂传播和复杂路由的结果，深入探讨复杂传染与简单传染的内在区别。

IV、主要参考资料：

- [1] Milgram S. The small world problem[J]. Psychology today, 1967, 2(1): 60-67.
- [2] Granovetter M. The strength of weak ties[J]. American journal of sociology, 1973, 78(6): 1.
- [3] Watts D J, Strogatz S H. Collective dynamics of ‘small-world’ networks[J]. nature, 1998, 393(6684): 440-442.
- [4] Kleinberg J. The small-world phenomenon: An algorithmic perspective[C]. Proceedings of the thirty-second annual ACM symposium on Theory of computing. ACM, 2000: 163-170.
- [5] Ghasemiesfeh G, Ebrahimi R, Gao J. Complex contagion and the weakness of long ties in social networks: revisited[C]. Proceedings of the fourteenth ACM conference on Electronic commerce. ACM, 2013: 507-524.

计算机 学院 计算机科学与技术 专业类 100611 班
学生 李强

毕业设计(论文)时间： 2014 年 03 月 04 日至 2014 年 06 月 01 日

答辩时间： 2014 年 06 月 09 日

成 绩： _____

指导教师： _____

兼职教师或答疑教师（并指出所负责部分）：

_____系（教研室）主任（签字）： _____



本人声明

我声明，本论文及其研究工作是由本人在导师指导下独立完成的，在完成论文时所利用的一切资料均已在参考文献中列出。

作者： 李强

签字：

时间： 2014 年 06 月



小世界网络中复杂传播和复杂路由的研究

学 生： 李强

指导教师： 任爱华

陈卫

摘 要

近年来的研究发现，现实世界中的许多网络都具有小世界性质，社会网络就是其中的代表。社会网络的小世界性质使得疾病、信息和谣言可以在社会网络上很快传播，这也引起了学术界的关注。深入理解信息在社会网络中的传播机制，以及如何控制疾病的传播、舆论的导向和营销策略的选择已经成为热点研究，影响力最大化问题和传播模型已成为许多领域涉及的重要内容。

复杂传染病（Complex contagion）是指网络的扩散机制，一个节点只有在其 $k \geq 2$ 个邻居都被感染后自己才会被感染。信息和病毒等简单传染病（Simple contagion）可以通过网络中的一条边进行扩散，而复杂传染病则要求多重接触。在 Kleinberg 的小世界网络模型中，强连接被认为是底层网格上确定的边，而弱连接指连接相距较远的节点之间的随机边。节点 u 和节点 v 通过一条弱连接相连的概率正比于 $1/|uv|^\alpha$ ，此处 $|uv|$ 是节点 u 和 v 之间的网格距离而 $\alpha \geq 0$ 是小世界网络模型的参数。近期的研究指出在弱连接可以极大地加快简单传染病的传播，而它在复杂传染病的传播中的作用并不显著。

本文首先研究了 Kleinberg 小世界网络中复杂传染病的传播（简称复杂传播），根据网络中弱连接的分布情况，定量地分析基于模型参数 α 的复杂传播时间，给出了 $\alpha < 2$ 或 $\alpha > 3$ 时传播时间的 n 的多项式形式的下界（ n 是网络中节点的个数），解决了以往遗留的开放性问题。

然后，本文类比 Kleinberg 的分散式路由，提出了基于复杂传染病的路由（简称复杂路由），同时对 Kleinberg 小世界网络上复杂路由时间进行了理论分析，求得复杂路由的路由时间在所有 α 范围内的 n 的多项式的下界。

最后，本文对比简单传染病和复杂传染病有关传播和路由的结果，说明了复杂路由要比复杂传播更难。同时这些结果也理论上支持了复杂传染病要比简单传染病更难传播的观点。

关键词： 计算社会科学，复杂传染，传播，分散式路由，小世界网络，社会网络



The Diffusion and Routing of Complex Contagion in Kleinberg's Small-World Networks

Author: Qiang Li

Tutor: Aihua Ren

Wei Chen

Abstract

Recent studies have suggested that small-world phenomenon is pervasive in real networks, including social networks. Diseases, information and rumors could spread fast in social networks exhibiting small world property, which have attracted wide academic attention. Deeply understanding the propagation mechanisms of social networks as well as controlling disease spread, guiding public opinion and marketing strategy have become heated research topics. Many recent works that involve many fields focused on the influence maximization in social networks and information propagation model.

Complex contagion refers to the propagation mechanism in a network where each node is activated only after $k \geq 2$ neighbors of the node are activated. Simple contagion, such as information and virus, could spread through a single contact, while complex contagion requires multiple activation. In Kleinberg's small-world network model, strong ties are modeled as deterministic edges in the underlying base grid and weak ties are modeled as random edges connecting remote nodes. The probability of connecting a node u with node v through a weak tie is proportional to $1/|uv|^\alpha$, where $|uv|$ is the grid distance between u and v and $\alpha \geq 0$ is the parameter of the model. Recent research shows that while weak ties can greatly speed up the diffusion of simple contagion, they are not as effective in the spread of complex contagion.

Firstly, this paper studies the diffusion of complex contagion (or *complex diffusion*) in Kleinberg's small-world networks. According to the probability distribution of weak ties, the author quantitatively analyses the diffusion time depending on the parameter of the parameter of the network model. Specifically, the author shows that for $\alpha < 2$ or $\alpha > 3$, the diffusion time is lower bounded by a polynomial in n (the number of nodes in the network), which addresses an open problem left last year.



Secondly, this paper investigates a new propagation phenomenon closer to decentralized routing proposed by Kleinberg, which is called the routing of complex contagion (or *complex routing*). Meanwhile the paper studies complex routing in Kleinberg's small-world networks and proves that routing time is lower bound by a polynomial in n for all range of α .

Finally, through comparing the results of diffusion and routing of both simple contagion and complex contagion, the author shows that complex routing is harder than complex diffusion. The results also provide theoretical support to the argument that complex contagion is much harder to propagate than simple contagion.

Key words: Computational social science, complex contagion, diffusion, decentralized routing, small-world networks, social networks



目 录

1 绪论	1
1.1 研究背景	1
1.1.1 小世界现象和六度分割理论	1
1.1.2 强连接和弱连接	2
1.1.3 小世界网络的建模	2
1.2 研究意义	3
1.3 国内外研究现状	4
1.3.1 简单传染病和复杂传染病模型	4
1.3.2 Kleinberg 模型的直径	5
1.3.3 其他相关工作	5
1.4 研究目标与内容	6
1.5 课题来源	7
1.6 论文的组织结构	7
2 相关技术综述	9
2.1 Kleinberg 小世界模型	9
2.2 复杂传染病模型	9
2.3 复杂传染病的传播	10
2.4 复杂传染病的路由	10
2.5 蒙特卡洛模拟	11
2.6 本章小结	11
3 蒙特卡洛模拟及实验结果分析	12
3.1 实验设计	12
3.2 实验环境	13
3.3 实验目的	14
3.4 实验步骤	14
3.5 实验结果及分析	14



3.6 本章小结	17
4 复杂传播的研究	18
4.1 复杂传播的主要结论	18
4.2 分析与证明	19
4.2.1 定理 4.1 (1) 的证明	19
4.2.2 证明方法框架	22
4.2.3 定理 4.1 (2),(3),(4) 的证明	24
4.3 讨论和扩展	29
4.4 本章小结	29
5 复杂路由的研究	30
5.1 复杂路由的结论	30
5.2 分析与证明	31
5.3 讨论和扩展	36
5.4 本章小结	36
6 总结与展望	37
6.1 工作总结	37
6.2 工作展望	38
致谢	40
参考文献	41



1 绪论

1.1 研究背景

社会网络（Social network）是由许多节点构成的一种社会结构。节点通常是组织或个人，节点之间的连接代表网络中个体之间的社会关系。社会关系可以是亲朋好友这种很亲密的关系，也可以是偶然相识的泛泛之交。这些社会关系，把社会网络上分散独立的个体连接起来，组成一个整体。

社会网络也被认为是传播疾病、信息和行为的媒介。社会学家研究社会网络和社会网络上的传播已经有几十年了，很多研究成果对于社会科学、经济学和计算机科学的交叉领域有很大的启示。这也不断激励着科学家们去对社会网络和社会网络上的传播建模。

1.1.1 小世界现象和六度分割理论

从图论的角度来看，小世界网络就是一个由大量节点构成的图，其中大部分节点之间的路径长度远小于图中节点数量。经常地，两个陌生人在交谈后发现有一个共同的朋友，互相不认识的人也许通过自己的朋友圈可以很快建立联系。那么人们所处的社会网络是不是小世界网络呢？

二十世纪 60 年代，美国哈佛大学社会心理学家斯坦利·米尔格伦（Stanley Milgram）做了一个连锁信实验^[1]。他将一些信件交给自愿的参加者，要求他们通过自己的朋友将信传到信封上指明的收信人手里。参与者们并不知道这个收信人详细资料，而是只知道收信人的基本信息，如所在的城市和职业。每个收到信件的人都被要求把信件转发给他们的朋友来尽可能地去把信件传递给指定的收件人。他发现，294 封信件中有 64 封最终送到了目标人物手中。在成功传递的信件中，平均只需要 5-6 次转发，就能够到达目标。这个实验表明，美国的任意两个人之间的“距离”是 6。这就是著名的“六度分隔”理论（Six Degrees of Separation）。尽管他的实验有不少缺陷，但这个现象引起了学界的注意。Milgram 的连锁信实验被称为是小世界实验，这开创了社会学研究的新领域，后续很多工作都在探索社会网络的小世界现象。

除了社会网络以外，小世界现象在自然形成的或者工业推动产生的网络中都是很普遍的。例如 C. 线虫的神经网络和美国西部的电力网^[2]，万维网中的超链接所构成的网络也具有小世界现象^[3]。



1.1.2 强连接和弱连接

马克·格拉诺维特(Mark Granovetter)在他最原始的工作中^[4,5]把社会关系分为强连接和弱连接。强连接(Strong ties)指非常亲密的关系,例如家人、挚友等。而且这些关系需要付出一定的时间和经历,需要用心去呵护。而相对于强连接,人们还维持着一些更广泛的社会关系,不需要投入太多心思,例如偶然结识的新朋友、远处的同学,这些关系被称为弱连接(Weak ties)。

Granovetter 通过调研麻省牛顿镇的居民如何找工作来探索社会网络^[5],得到了一个令人惊讶的结论。被调查的人中,大部分都是通过自己的人脉来找到现在的工作但是比较有趣的是这些人脉大部分都不是最亲密的那些朋友,这被称为弱连接的力量^[4](The strength of weak ties)。这也许是因为一个人的比较亲密的朋友他们可能本来已经是朋友,这些社会关系比较冗余,对于信息的远距离传递帮助没有那么大。而少数的弱连接,却可以帮助信息传递很远,对于社会网络中信息的传播有很大的帮助。

1.1.3 小世界网络的建模

社会网络小世界现象的发现和弱连接理论的提出引起了学术界的关注,许多工作开始专注于小世界网络的建模。Watts 和 Strogatz 首先提出了基于圆环的小世界网络模型^[2],模型按照如下步骤构造。

- 1、底层框架是一个圆环,由 n 个节点的集合 V 均匀分布在圆环上。
- 2、把所有节点与其最近的 k 个邻居相连,这里 k 是一个较小的常数,这些边被称作强连接。
- 3、对于这个网络中每条强连接,以 p 的概率重新选择边的终点,终点是完全随机地从 V 中选择的,重新连接的边被称作弱连接。

这样就完成了小世界网络的建立。他们还提出了小世界网络的应该具有的两个特性:直径较小(Short diameter)和聚集系数较大(High clustering coefficient)。直径较小是因为小世界网络中两个不认识的个体可以找到一条较短的连接两个人的路径,这首先要网络中任意两点之间存在比较短的路径。聚集系数是一个节点的邻接点之间相互连接的程度,真实社会网络中,一个人的朋友之间一般都会相互认识,这说明社会网络的聚集系数也比较大。

之前的研究^[6]一般设定每个人都完全随机地从所有人选择一些人作为自己的朋友,而且相识关系是对称的。这样构成的网络是一个随机图,而且的确有比较小的直径^[7]。但是过于随机的网络导致这个网络的聚集系数比较小,这不符合社会网络的特性。然



而, 如果一个网络的聚集系数比较大, 例如一个二维网格, 这个网络的直径会较大, 没有小世界现象的存在。Watts 和 Strogatz 提出的这个模型是介于两个极端之间的模型, 同时具有较小的直径和较大的聚集系数。这个模型是一个圆环框架和随机图的叠加, 圆环作为模型的底层结构 (Underlying structure), 而随机图代表网络中少量的弱连接。这样叠加图的模型恰好同时保持了圆环较大的聚集系数和随机图较小直径的优点。由于聚集系数较大, 信息在一个聚集的群体中通过强连接可以很快传播。而为数不多的弱连接可以把消息传递到很远的群体, 进而消息可以在整个网络中很快传播。

Kleinberg 改进了 Watts 和 Strogatz 的模型^[8], 也是基于由强连接构成的框架和随机图的弱连接叠加而成。Kleinberg 的小世界网络的框架是一个二维的网格, 网格上的边代表强连接。同时网格上的每个节点 u 都会发出一条弱连接, 而这条弱连接以 v 为终点的概率正比于 $1/|uv|^\alpha$, 这里 $|uv|$ 是节点 u 和节点 v 之间的网格距离, α 是小世界模型的参数 (Clustering exponent)。Kleinberg 证明了当且仅当 α 等于网格的维度 (这里是 2) 时, 贪心的分散式路由算法可以有效地把消息传递给指定的目标^[8]。这里的分散式路由算法是因为在算法执行的每一步, 算法只知道所有已经传递过消息的节点的邻居, 而不清楚整个网络中弱连接的分布, 可以理解为没有路由表。而算法的贪心则是指在每一步, 当前节点会选择距离目标网格距离最近的节点来传递消息。 $\alpha = 2$ 的小世界网络中分散式路由的效率在数量级上符合 Milgram 的实验结果。然而在 α 不等于网格的维度时, 却不存在有效的路由算法, 路由时间出现了从对数级到多项式级的相变。特别的, Newman 和 Watts 提出的小世界模型^[9] 恰好是 Kleinberg 的模型在 $\alpha = 0$ 的一维形式。在 $\alpha < 2$ 时, 弱连接的分布比较随机, 相对来说会更易于连接到相距较远的点。但是弱连接分布的过于随机化导致在分散式路由算法执行时, 个体不知道应该把消息传递给哪一个联系人, 不能保证路由算法的高效率。随着 α 的增大, u 的弱连接的分布逐渐集中在 u 的附近, 弱连接的分布更依赖地理位置。当而 $\alpha > 2$ 时, 节点弱连接的分布就过于集中, 聚集系数较大。因此即使有弱连接的帮助, 分散式算法也不能有效的找到目标。而在 $\alpha = 2$ 这个位置, 弱连接不会过于集中, 同时也具有一定的地理依赖性, 弱连接的随机性和地理性取得了最好的权衡, 分散式算法也取得了最好的效率。

在本文中, 我们使用的就是 Kleinberg 的小世界模型 (Kleinberg's small-world network model)。

1.2 研究意义

现实中大量的网络都具有小世界现象, 从互联网到万维网, 从神经网络到电力传输网络等。人类社会的网络化, 一方面可以提供更高效高质量的生活, 提升社会生产



力。但是同时,过于网络化的社会有时也会带来灾难。例如传染病的爆发,到处肆虐的计算机病毒,谣言的扩散,网络的小世界性质为这些有害信息的传播提供了良好的条件。同时网络中节点的互相影响也会因其连锁反应,例如电力网络的瘫痪。人们必须了解传播的机制,更好的理解网络的动力学特征和结构之间的关系,才能有效地控制和预防这些灾难的发生。对于社会网络复杂传染病传播和路由的研究可以帮助理解小世界网络的结构和复杂传播速度的关系,对于疾病预防和谣言控制有一定的帮助。复杂路由的相关研究则对社交网站的好友推荐和主动交友有一定的启示。

1.3 国内外研究现状

1.3.1 简单传染病和复杂传染病模型

Granovetter 在 1978 年的另一篇文章^[10]中提出了有阈值的传染模型用来描述谣言、创意和暴动行为等的传播,这些传播往往需要接受者投入一定的成本。在这个传染模型中,社会网络中的个体只有在他的感染疾病邻居的数量超过一定阈值的时候才会被传染。这个阈值模型启发了 Kempe 等人,他们在 2003 年提出了线性阈值,固定阈值和通用阈值等模型^[11],这与我们研究的复杂传染病模型直接相关。

在近期的工作中,Centola 和 Macy 把阈值传染模型归类为简单传染病模型和复杂传染病模型^[12]。简单传染病对应于网络中所有节点阈值均为 1 的阈值传染模型,这就意味着当一个节点在有一个被感染的邻居的时候他就会被感染。简单传染病模型在现实生活中可以看作流感或者八卦的传播,因为人们在接触到消息源或者传染源后很容易就被影响。因此,一条弱连接就可以把消息传递到很远的区域,这极大地加速了消息的传播。而另一方面复杂感染病对应于阈值至少为 2 的阈值传染模型。因此个体只有在他的一定数量的邻居都被感染的时候才会被感染。对应到实际的社会网络,复杂传染病模型更多地描述的是那些有一定代价的决定的传播,譬如购买一款新产品、采纳一个有风险的建议。一款新式手机的上市后,一个人往往会在很多同学和同事都购买了该款手机后才会动心购买,新电影的上映也是如此。个体需要在接收到来自足够多独立渠道的建议后才能说服自己作出决定。Centola 和 Macy 进一步指出弱连接尽管在社会网络中对于信息的长距离传输有很大帮助,但是对于复杂传染的帮助比较小。这是因为对于社会网络中的复杂传染病模型,只有在不同区域之间形成较长和较宽的由弱连接组成的桥梁时,信息才能较快的传播。这对于随机性比较大的而数量较少弱连接来说是很难形成的。



1.3.2 Kleinberg 模型的直径

在 2000 年 Kleinberg 提出了一个小世界模型并讨论了在这个模型上分散式路由的效率。但是他只是讨论了路由算法的效率,并没有去分析网络的直径。Nguyen 和 Martel 在最近的工作中讨论了当模型参数 $0 \leq \alpha \leq 2$ 时 Kleinberg 模型的直径^[13]。他们的结果是 $0 \leq \alpha \leq 2$ 的 Kleinberg 网络模型的直径为 $\Theta(\log n)$, n 是网络中节点的个数。而在每个节点的度为常数时,一个网络直径的期望也是 $\Omega(\log n)$,这说明 Kleinberg 模型的直径的确很小,在 $0 \leq \alpha < 2$ 时分散式路由的效率较低不是因为网络两个节点之间不存在较短的路径,而是算法不能找到这些路径。这也从侧面说明直径和路由的不同。如果把时间看成离散的时间点,传播是指给定了种子节点后,在每一个时间点,在传染模型下当前所有可以被感染的节点都会被感染。而路由则是在每个时间点,从当前可以被感染的节点集合中,根据路由算法仅选择一个节点感染。一个网络的直径恰好和传播速度对应着,从这我们就知道了对于同一个网络,传播速度和路由速度可能相差很大(n 的对数级和多项式级的差距)。

接下来 Nguyen 和 Martel 又证明了 $2 < \alpha < 4$ 时, Kleinberg 基于二维网格的模型也具有 $O(\log^\beta n)$ 形式的直径, β 是和 n 无关的常数^[14]。而随着 α 继续增大,在大于 4 的时候, Kleinberg 模型的直径就变成了多项式形式,出现了从对数到多项式的相变现象。

1.3.3 其他相关工作

受到有关复杂传染模型工作的启发, Ghasemiesfeh 等人首次给出了小世界模型中复杂传染的理论分析^[15]。他们研究了 k -复杂传染病的传播(简称 k -复杂传播),在这个设置下网络中每个节点的阈值为 k ,也就是说节点在有 k 个被感染邻居时就立即会被感染。他们研究了传播时间,也就是给定 k 个通过强连接相连的种子节点后,网络中所有节点都被感染所需要的时间。他们的主要结果是在 $k = 2$ 对于基于二维网格的 Kleinberg 小世界网络上,每个节点发出两条无向的弱连接,同时网格距离在 2 以内的点会被强连接相连。这时,传播时间具有上界 $O(\log^{3.5} n)$, n 仍然指小世界网络的节点个数。同时也给出了基于一维圆环 Newman-Watts 模型^[9]的结论,对应 $\alpha = 0$ 条件下的 Kleinberg 模型。这时弱连接是完全均匀分布的,而此时传播时间具有下界 $\Omega(n^{\frac{1}{3}})$ 。然而,他们留下了一系列尚未解决的问题,这也是本文的动机所在。

自从上文提到的 Watts-Strogatz 模型^[2]和 Kleinberg 模型^[8]被提出后,还有其他很多小世界模型的扩展和变种被研究过。在 2001 年 Kleinberg 又提出了基于树状结构的小世界模型,树的叶子代表社会网络中的个体,然后根据个体之间最小公共子树的深



度来添加随机边。可以证明当每个节点发出的边的个数为 $\Theta(\log^2 n)$ 时, 贪心路由算法可以很快地找到目标。Fraigniaud 和 Giakkoupis 把 Kleinberg 的小世界模型和幂律分布 (Power law) 结合起来^[16], 即每个节点发出的弱连接数目遵从幂律分布。同时他们也给出了由任意的底层网络和少数弱连接结合成的网络中路由算法的分析^[17]。

在 Kempe 提出了影响力传播最大化的问题之后^[11, 18], 一系列后续问题都在研究在随机传播模型下给定限额 *budget* 后, 如何找到使社会网络中最终被影响的节点最多的 *budget* 个种子节点。这实际上是一个 NP 问题, 最近工作都在研究比较高效的启发式算法。陈卫等人提供了一个可以在大规模网络上运行的影响力最大化算法^[19, 20], 在有较高的效率的同时也可以输出不错的结果。与此同时陈宁证明了在一个所有节点都有固定阈值的网络中, 近似算法很难找到可以影响到所有节点的最小种子节点集合^[21], 也就是说很难给出一个多项式对数近似比 (Polylogarithmic factor) 的近似解。

1.4 研究目标与内容

现有的有关小世界网络的分析大多是关于简单传染病的传播和路由, 去年 Ghasemiesfeh 等人给出了第一份有关复杂传染病的传播的理论分析^[15]。本文进一步给出 Kleinberg 模型下复杂传染病的传播理论上的分析, 提升传播时间的下界, 解决 Ghasemiesfeh 等人留下的开放性问题。同时在论文准备研究一种新的消息传播行为, 复杂传染病的路由, 这有些像 Kleinberg 的分散式路由, 而在本文关注的是复杂传染病上的分散式路由。最后对比简单传播, 简单路由, 复杂传播和复杂路由的效率。

我们的研究内容大致制定如下:

第一, 进行蒙特卡洛模拟, 随机生成 Kleinberg 小世界网络, 在生成的小世界网络上估计简单传染病的传播时间和路由时间以及复杂传染病的传播时间, 然后分析实验结果。

第二, 给出复杂传染病传播的理论上的分析。在 2013 年 Ghasemiesfeh 等人第一次给出了复杂传染病模型下信息传播速度的研究。他们同时也证明了在小世界模型的参数 α 为网格维度时, 复杂传播的速度很快, 是网络中节点数量的多项式对数级 (Polylogarithmic)。而对于参数 α 不等于网格维度时, 他们只给出了很弱的传播时间的下界。本文主要关注对于二维网格模型, 当 α 不为 2 时, 复杂传播时间的下界, 尝试去提升已有的下界。希望可以得到类似于 Kleinberg 的分散式路由算法的结果, 当且仅当在 α 等于 2 时, 复杂传播时间较短, 当 α 不为 2 时, 可以给出节点数量多项式形式的下界。

第三, 研究一种新的消息传播行为, 复杂传染病的路由。之前的研究已经给出了



简单传染病的路由（Kleinberg 的贪心路由）、简单传染病的传播（网络的直径）和复杂传染病的传播时间上下界。但是复杂传染病的路由一直没有人进行相关研究。相比于简单传染病的路由，复杂传染病的路由需要更多的弱连接来支持。复杂路由所需要的时间必然要比简单路由更多，但是仍然希望看到类似于简单路由的在 α 等于网络维度时具有最高的效率。在文中也会去寻找对于复杂传染病的比较高效的路由算法。复杂路由在现实生活中类似于一群人在不断的扩大自己阵营的同时去影响另一个人。复杂路由的研究对于社交网站上主动交友方面也有一定的帮助^[22]，即在人人网和 Facebook 等社交网站上通过添加一些中间好友来增加目标接受自己好友请求的概率。

第四，分析对比简单路由、简单传播、复杂路由和复杂传播的结果，研究传播时间和路由时间与模型参数 α 的关系，同时深入探讨复杂传染与简单传染的内在区别。

1.5 课题来源

本毕业设计课题来源于与微软亚洲研究院的陈卫研究员和中科院计算所的孙晓明老师、张家琳老师合作完成的论文“The Diffusion and Routing of Complex Contagion in Kleinberg’s Small-World Networks”。

1.6 论文的组织结构

本文主要是研究 Kleinberg 小世界网络中复杂传染病的传播时间和路由时间，首先分析了小世界模型的特性，包括已有的关于简单传播（直径）和简单路由的结论。然后，对相关小世界模型和复杂传染病模型的定义做出详细论述。其次，给出了有关小世界网络上传播和路由的蒙特卡洛模拟实验的完整过程，包括设计、步骤和结果分析等。接下来给出有关复杂传播和复杂的主要结论以及完整证明。最后给出本文结论和已有结论的分析对比。

论文组织结构如下：

第一章绪论

介绍课题的研究背景，阐述了小世界现象的研究背景和以及国内外研究现状，提出本文的研究目标和主要工作内容。

第二章相关技术综述

介绍本文工作涉及到的主要相关理论及模型，重点分析了 Kleinberg 小世界网络以及复杂传染病的相关知识，然后介绍了蒙特卡洛模拟的相关技术。

第三章蒙特卡洛模拟及实验结果分析

首先给出了基于 Kleinberg 小世界网络的蒙特卡洛模拟实验的设计，然后介绍实验



目的和步骤，最后给出实验结果和分析过程。

第四章复杂传播的研究

首先给出复杂路由的主要结论，然后给出证明的框架和基于框架给出完整的证明。最后把结论扩展到 k -复杂传播并与简单传播进行比较。

第五章复杂路由的研究

给出复杂路由的结论及完整的证明过程，然后探讨一步允许感染 m 个节点的复杂路由，建立起复杂传播和复杂路由的关系。

第六章总结与展望

在最后的总结与展望中，首先整体给出了文章的结论，然后对比本文结果和已有结果，深入讨论了复杂传播和简单传播的内在差别。然后指出已有工作的不足，并对下一步的工作进行展望。



2 相关技术综述

本章主要介绍本文中使用的 Kleinberg 小世界网络的精确定义, 包括强连接和弱连接的生成方式。同时也介绍文中用到的简单传染病和复杂传染病模型以及两种信息扩散的方式: 传播和路由。最后介绍在模拟小世界网络上的传播和路由时采用的蒙特卡洛模拟技术。

2.1 Kleinberg 小世界模型

Kleinberg 的小世界模型是一个由 n 个节点的集合 V 生成的随机图。 n 个节点分布在一个 $\sqrt{n} \times \sqrt{n}$ 的二维网格上^[8], 为了方便, 我们把网格的上边界和下边界连接起来, 同时也把网格的左边界和右边界连接起来。这样二维网格就变成了“环面”, 网格中每个节点的位置都是对称的。网格上两个节点 u 和 v 之间的曼哈顿距离 (Manhattan distance) $|uv|$ 是在网格上 u 到 v 的最短路径的长度。

这个随机图上有两种类型的边: 强连接和弱连接。强连接是任意两个曼哈顿距离不超过 p 的节点之间生成的无向边, 这里 $p \geq 1$ 是一个模型的常数。弱连接指连接节点 u 和网格上可能相距较远的节点 v 之间的随机边。每个节点 u 会有 q 条互相独立的弱连接边, u 的第 i 条弱连接边以 v 为终点的概率正比于 $1/|uv|^\alpha$, $\alpha \geq 0$ 是小世界模型的参数。我们用 $1/|uv|^\alpha$ 乘以归一化因子 $\mathcal{Z} = 1/\sum_{v \in V} |uv|^{-\alpha}$ (在环形网格上, 这个值对于任意的节点 $u \in V$ 都相等), 这样就得到了弱连接的概率分布函数。最初 Kleinberg 描述的网络模型^[8] 中, u 到 v 之间的弱连接被认为是有向边, 这样的网络被称为有向 Kleinberg 小世界网络模型。而有些研究工作^[15] 中弱连接被认为是无向的, 这样的网络被称为无向 Kleinberg 小世界网络模型。两个模型在本文中都被讨论了, 在分析复杂传染病的传播时, 我们为了和以前的工作保持一致, 采用无向 Kleinberg 小世界网络模型。分析复杂传染病的路由时, 我们使用有向 Kleinberg 小世界网络模型。

2.2 复杂传染病模型

本文用传染病去模拟信息、疾病和想法在网络中的扩散。网络中的节点都有三种状态: 未感染 (inactive)、暴露 (exposed) 和已感染 (activated)。节点可以从未感染状态转变为暴露状态, 然后再进入感染状态, 但是不能反方向转变, 例如不能从已感染的状态变成未感染状态。

传染病传染的过程可以用离散的时间步骤 $0, 1, 2, \dots$ 来描述。如果一个节点 u 在



$t-1$ 时间有至少 k 个已经感染的邻居节点（在有向图中指有 k 个已感染节点发出的有向边指向 u ），那么在时间 $t \geq 1$ 节点 u 就变为暴露状态， k 是传染病的阈值。一个已经暴露的节点可以立即或者在后续时间步骤中转变为感染状态，这取决于消息扩散的方式。简单传染病指 $k=1$ 的传染病，也就是说 u 的一个已感染的邻居就可以让 u 变为感染状态（有可能感染 u ）。而复杂传染病则较难传染，因为复杂传染病的阈值 $k \geq 2$ ，要感染一个新的节点，至少需要两个已感染的邻居节点。阈值 $k \geq 2$ 的传染病称为 k -复杂传染病。

2.3 复杂传染病的传播

在本文中，我们定义复杂传染病的传播（简称复杂传播）为在传染病扩散过程中，所有暴露状态的节点在同一时间会被立即感染。

在研究 k -复杂传播时，为了保证网络中所有节点都会被感染，最初选取在网络上 k 个连续的节点设为感染状态，这些节点被称作种子节点。为了方便，本文设置 $p=q=k$ ，这样每个节点 u 会发出 k 条弱连接，而 u 会与曼哈顿距离在 k 以内的节点建立强连接。当 $p=k$ 时， k -复杂传播仅仅通过强连接也可以感染整个网络， $q=k$ 也可以让一个节点的弱连接提供传播的机会。

本文主要研究传染病多快把整个网络都传染，这个可以用传播时间来描述，而在网络中的强连接和弱连接都固定后传播时间是定值。传播时间是指从 k 个种子节点开始，整个网络都被感染时经过的时间步数。

2.4 复杂传染病的路由

本文进一步研究了一种比较像分散式路由^[8]的扩散方式，我们称之为复杂传播病的路由（简称复杂路由）。

在复杂路由中，最初会选择一个节点 t 作为目标，同时也有 k 个种子节点（和复杂传播中设定一致）。复杂路由的任务是尽快感染到目标节点 t ，不同于复杂传播的是，每一个时间步骤中所有被暴露的节点不会被立即感染。每一步，我们只能从暴露状态的节点集合中选择一个节点来感染。选择节点的策略称为路由策略。更进一步，当在时间 i 选择了节点 u 去感染时，算法仅仅知道已感染节点集合的所有邻居（在有向图中是已感染节点发出的有向边指向的节点），这就是分散式路由的原则。复杂路由在日常生活中的情景比较像一群人想要去影响一个目标，但是无法直接影响到目标。于是这群人逐渐的扩大自己的势力，拉拢他们认为可以对劝说目标有帮助的人入伙，最终影响到目标。在扩大势力的时候，他们只知道已经入伙的人的朋友，而且拉拢入伙需要



一定的代价, 在每一步只能选择拉拢一个人。注意到如果把 k -复杂传染病用简单传染病替代, 并且要求下一个被感染的节点是最新感染节点的邻居, 这就是 Kleinberg 研究的分散式路由^[8]。

为了研究路由找到目标的速度, 本文定义路由时间为通过路由方式感染距离种子节点曼哈顿距离最远的目标节点 t 所需要的步骤。

2.5 蒙特卡洛模拟

蒙特卡洛模拟 (Monte Carlo simulation), 也称统计模拟方法, 是二十世纪四十年代中期由于科学技术的发展和电子计算机的发明, 而被提出的一种以概率统计理论为指导的一类非常重要的数值计算方法。是指使用随机数 (或更常见的伪随机数) 来解决很多计算问题的方法, 与此对应的是确定性算法。有些问题很难直接求解, 例如一些期望、积分等, 直接求出精确解可能时间复杂度很大。这时我们就可以采用蒙特卡洛模拟。主要有两部分工作:

- 1、用蒙特卡洛方法模拟某一过程时, 需要产生各种概率分布的随机变量。
- 2、用统计方法把模型的数字特征估计出来, 从而得到实际问题的数值解。

例如在分析复杂传播时间的期望时, 直接根据小世界网络中弱连接的分布函数求出传播时间的期望比较难。可以根据弱连接的概率分布函数来生成弱连接, 在给定的生成的小世界网络中, 可以计算出复杂传播的时间, 这一步就是产生符合概率分布的随机变量。做大量的独立重复的生成小世界网络的实验, 利用实验中记录的传播时间的平均时间作为传播时间的期望。这一步是通过统计得到数值解。

2.6 本章小结

本章主要介绍了本文使用的 Kleinberg 小世界模型以及传染病模型和传播路由等概念, 并给出了精确的数学定义, 同时提到了蒙特卡洛模拟。2.1 节介绍了 Kleinberg 的小世界模型, 主要分析了弱连接的生成方式。2.2 节介绍了复杂传染病模型, 终点介绍阈值和节点的三种状态。2.3 节介绍了复杂传染病的传播, 2.4 节介绍了复杂传染病的路由并与简单路由做了一些对比。2.5 节介绍了蒙特卡洛模拟, 简要提及了本文中复杂传播的模拟方法。



3 蒙特卡洛模拟及实验结果分析

在进行理论分析工作之前,首先准备进行蒙特卡洛模拟实验,通过蒙特卡洛模拟,得出传播时间和路由时间与网络大小以及 Kleinberg 小世界模型参数 α 的关系,有利于进一步的理论分析。根据实验结果,也可以更好的理解简单路由在 $\alpha = 2$ 时路由时间发生的相变现象,通过数据去理解已有的成果并为后面的理论分析打下基础。

3.1 实验设计

首先需要生成 Kleinberg 的小世界网络,这就需要设计如何存储和生成一个规模较大的随机图。与此同时,需要考虑如何生成遵从逆 α 次方分布的弱连接。

比较好的是,二维网格上的节点都是平等的,每个节点的弱连接的分布都是一致的。这样对于网络上的一个节点 u ,弱连接的生成可以按照如下步骤:

- 1、提前按照节点 u 的一条弱连接以 v 为终点的概率 $\frac{|uv|^{-\alpha}}{\sum_{w \in V} |uw|^{-\alpha}}$ 生成概率分布函数。因为节点是离散的,可以用数组来存放概率分布函数,数组下标对应着节点的编号。
- 2、对于节点 u 的每一条弱连接,首先随机生成一个 $[0, 1]$ 之间的随机数 p ,然后在概率分布函数的数组上面进行二分查找,找到值 p 对应的数组下标 i_v 。则对于 u 的这条弱连接,终点即为下标 i_v 对应的节点 v 。

用上面的方法就可以生成对于一个节点 u 的弱连接,生成一张图中所有节点的弱连接就可以得到一个完整的 Kleinberg 小世界网络了。下面讨论在生成的小世界网络上的传播和路由的实现。

简单传播的过程就是广度优先搜索的过程,因此用广度优先搜索遍历网络,计算最后被遍历的节点需要多少轮广度搜索,这就是简单传播的步数。简单传播的步数是从种子节点到网络中相距最远节点的路径的长度,也就对应着网络的直径。对于复杂传播(考虑 2-复杂传播),也比较类似,可以为每个节点设置一个状态,然后用类似广度优先搜索的方法开始遍历网络中的节点,大致算法如下:

- 1、生成 Kleinberg 小世界网络,建立一个遍历队列,种子节点入队。每个节点有个状态变量 $state$ 用来记录已感染的邻居节点的数量。初始状态时所有节点的 $state$ 都为 0。节点用 $step$ 变量记录被感染的时间步骤,种子节点的 $step = 1$ 。
- 2、队头出列,队头元素对应的节点为 u ,对于 u 的每个没有被感染的邻居 v ,如果 v 的 $state = 0$,则令 $state = 1$;如果 v 的 $state = 1$,则令节点 v 入列,同时



$step_v = step_u + 1$, 标记 v 为被感染状态。

3、 重复步骤 2 直到网络中所有节点都被感染或者队列为空。

这样遍历完整个网络之后, 最后一个出列的节点的 $step$ 就可以看做传播时间。从复杂传播的模拟算法中可以看出, 首先需要生成整个 Kleinberg 小世界网络, 即为网络中的每个节点都生成弱连接。同时至少需要遍历网络中的所有节点, 因此模拟时比较耗费时间, 简单传播的模拟也类似。为了存储生成的小世界网络, 采用了邻接链表的方法, 比较节省空间。

而简单路由的模拟则比较简单, 因为是分散式路由算法, 算法不知道除了已感染节点之外节点的弱连接, 模拟过程也严格遵从延迟选择原则。模拟算法步骤如下:

- 1、 在网络上选取种子节点 s 和目标节点 t 。设 $holder$ 为当前消息持有者, 初始状态 $holder = s$ 。
- 2、 根据概率分布函数生成节点 $holder$ 的弱连接, 然后从 $holder$ 的邻居节点 (通过强连接和弱连接相连的邻居) 中选择距离 t 曼哈顿距离最近的节点 v , 令 $holder = v$ 。
- 3、 如果 $holder$ 为 t , 则模拟结束, 否则重复步骤 2。

步骤 2 的执行次数就是简单路由需要的步数。可以看到简单路由在模拟过程中, 不需要生成整个小世界网络, 只需要在消息到达节点 $holder$ 时动态生成节点 $holder$ 的弱连接即可。而简单路由的步数本来就比较少, 所以模拟时只需要生成路由路径上节点的弱连接, 效率很高。

3.2 实验环境

因为实验需要统计的数据是模拟程序的输出, 所以实验结果与硬件无关, 但是在这里还是做简要的介绍。选用微软亚洲研究院的服务器作为实验主机, 相关参数如下:

1) 硬件环境

- CPU: Intel(R) Xeon(R) CPU E5330 @ 2.40 GHz(2 processors 16 Core)
- 内存: 48GB

2) 软件环境

- 操作系统: Windows Server 2008 R2 Enterprise
- IDE: Microsoft Visual Studio 2010
- 绘图工具: gnuplot 4.6.4, MATLAB 7.11.0(R2010b)



3.3 实验目的

本实验的目的是通过蒙特卡洛模拟, 得到在小世界网络大小 n 和模型参数 α 时简单传播时间、简单路由时间和复杂传播时间的期望的数据。然后可视化数据, 清晰地观察 n, α 对传播时间和路由时间的影响, 进一步理解已有的成果, 也帮助我们去预测估计复杂传播的结论。

3.4 实验步骤

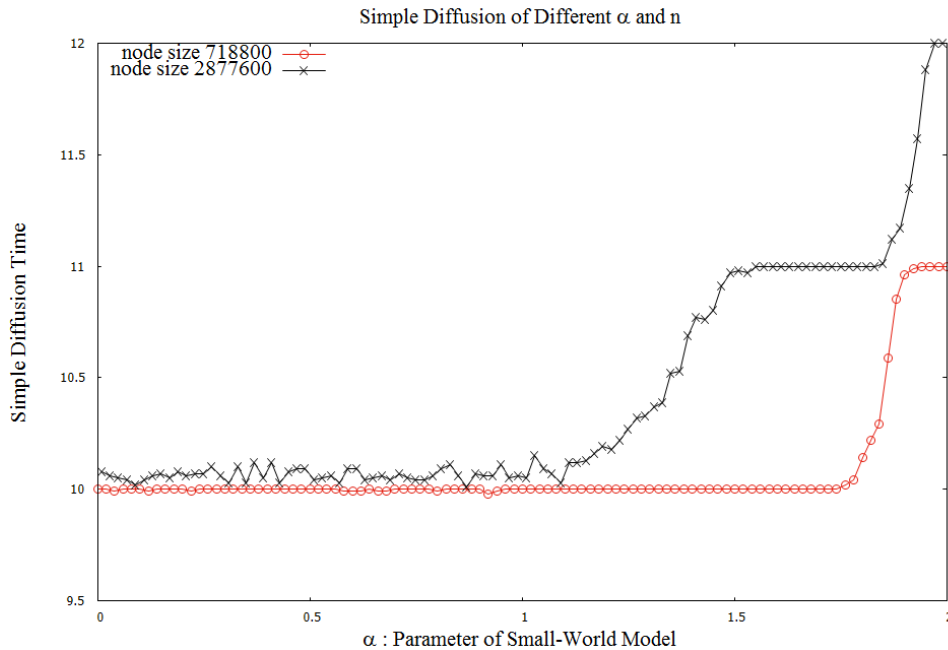
对于每个 n 和 α , 进行 1 万次模拟, 统计平均结果, 作为期望传播时间或者路由时间。对于简单传播, 模型参数 α 取值范围为 $[0, 2]$, 步长为 0.01。网络节点个数 n 取值为 718800 和 2877600, 因为需要生成整个小世界网络, 模拟效率比较低, 选用较大的网络时模拟程序运行时间过长。简单路由由于效率较高, 模拟时可以选用比较大的网络, 因此对简单路由进行模拟时, 网络节点个数 n 取值为 10^6 、 2.5×10^7 、 10^8 、 4×10^8 、 9×10^8 、 2.5×10^9 和 3.6×10^9 。复杂传播的效率较低, 仍然采用比较小的网络进行模拟, 网络节点个数取值为 217140、433380、867244 和 1731660。网络节点选取的数量大致构成以 2 为公比的等比数列, 这样可以更明显地观察出结果的量级是否是对数。为了方便后续的理论分析, 复杂传播模拟时 α 的取值范围为 $[0, 4]$ 。

模拟完成后会生成列表格式的数据, 第一列为模型参数 α 的取值, 第二列为网络节点 n 的大小, 第三列为模拟的传播时间或者路由时间。把分别对应简单传播, 简单路由和复杂传播的数据用 gnuplot 可视化。

3.5 实验结果及分析

本节对实验数据进行分析, 结合可视化后的数据, 深入理解简单传播、简单路由和复杂传播。首先是简单传播的传播时间随网络大小 n 和 Kleinberg 小世界模型参数 α 的变化曲线, 如图 3.1 所示。

图的横坐标为模型参数 α 的值, 纵坐标是简单传播所需要的步数。图上的两条曲线分别对应着节点个数为 718800 和 2877600 的小世界网络的模拟情况。简单传播时间是用广度优先搜索来求出的, 最后一个被感染的节点需要广度优先搜索的轮数被记为是简单传播时间。然后对于每个 n 和 α 模拟多次后结果取平均值得到期望的简单传播时间。从图 3.1 可以看出简单传播时间随着网络的增大也在增大, 这是因为网络的直径随着网络节点的增加必然会增大。但是增大的趋势比较小, 这也与理论分析的在 $\alpha \in [0, 2]$ 时具有 n 个节点的网络的直径为 $\Theta(\log n)$ 相符合。而简单传播时间随着 α 的

图 3.1 简单传播时间随 n 和 α 的变化曲线

减小也在减小。 α 越小，网络中弱连接的分布就越随机。随着 α 的增大，弱连接的分布开始趋于聚集，网络的直径开始增大。而在 $\alpha = 0$ 时，网络中的弱连接是均匀分布的，由随机图的理论^[7] 可知此时网络的直径为 $O(\log n)$ 。对简单传播图像的分析也知道 Kleinberg 的小世界网络的确直径比较小，两个点之间的确被比较短的路径相连。

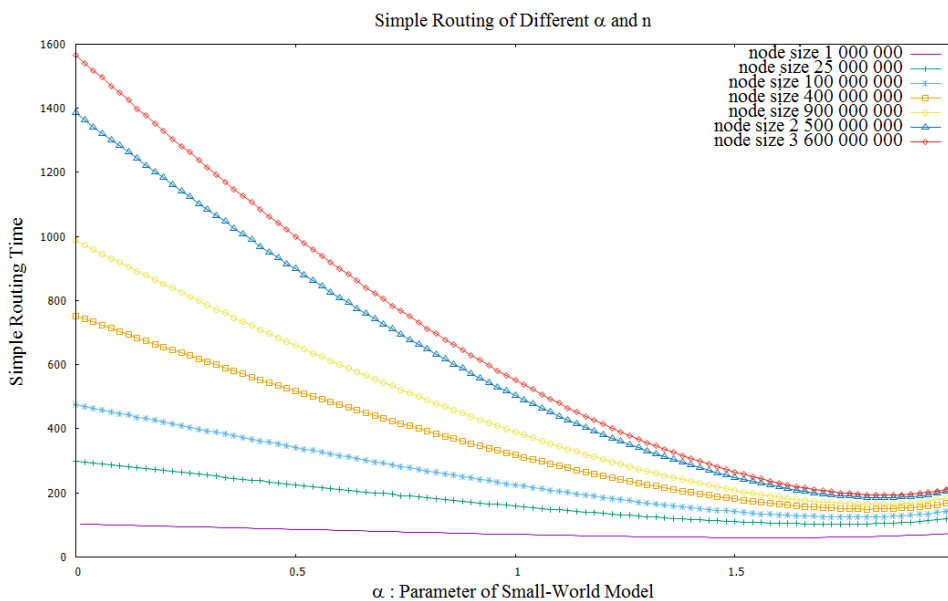
图 3.2 简单路由时间随 n 和 α 的变化曲线

图 3.2 给出了简单路由时间和 Kleinberg 小世界模型的关系。图的纵坐标为简单路由需要的步数，也就是简单路由时间，横坐标是取值区间为 $[0, 2]$ 的 α 。可以看到大约

是在 $\alpha = 2$ 的附近, 简单路由的效率最高, 但是和理论解释的在 $\alpha = 2$ 的点取得最小值不符。仔细观察可以发现当 n 比较大时, 路由步数最低点对应的 α 逐渐向 2 偏移。可以预测, 当 n 足够大时, 简单路由会在 $\alpha = 2$ 处取得最高的效率, 这就和理论相符了。同时也可以看到, 简单路由的效率的确很高, 在 36 亿个节点的网络中, 路由算法也可以在 1600 步以内找到目标。随着 α 的减小, 路由的步数反而开始增加, 这和网络直径随着 α 的减小而减小相反。这也说明简单路由不仅仅是网络中两点之间存在比较短的路径, 而是弱连接的分布可以帮助分散式算法找到这些比较短的路径。尽管 α 较小时, 网络的直径比较小。但是弱连接的分布过于随机, 分散式算法无法找到这些短路径。再对比简单传播和简单路由的速度。在网络中节点个数是百万量级时, 简单传播大约只需要十多步就可以感染整个网络, 而简单路由仍然需要大约 100 步才能找到目标。这也说明分散式路由虽然可以较快的找到目标, 但是找到目标所需要的步数还是比较大的。通过分散式路由找到的到达目标的路径和两个节点之间的最短路径还是有很大差别的。

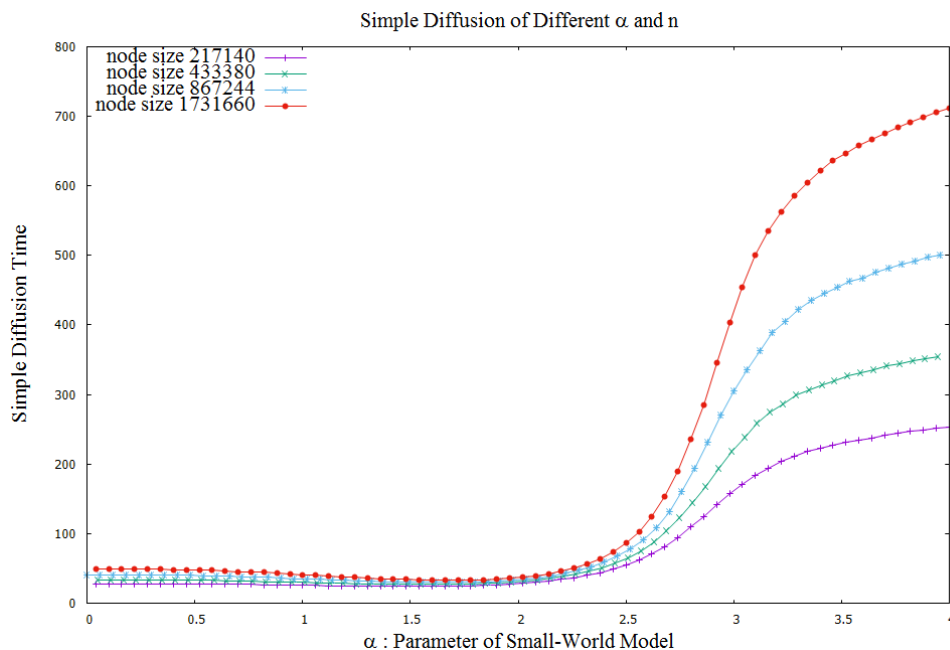


图 3.3 复杂传播时间随 n 和 α 的变化曲线

图 3.3 给出了复杂传播时间随着网络节点个数和模型参数 α 的变化关系。首先可以看到, 复杂传播在 $\alpha = 2$ 的附近时, 传播时间最少。Kleinberg 的分散式路由也是在 $\alpha = 2$ 处取得最少的路由时间, 所以我们期望在复杂传播中也可以看到类似的结果。根据可视化的数据, 复杂传播的确也是在 α 附近时传播时间最少。进一步可以观察在 $\alpha > 2$ 时, 传播时间明显上升, 和 $\alpha \leq 2$ 时的传播时间相差很大。去年的结果^[15] 显示在



$\alpha = 2$ 时, 复杂传播时间上界为 $O(\log^{3.5} n)$ 。因为观察到复杂传播在 $\alpha = 2$ 时效率最高的现象, 后续工作更多地集中在给出其他 α 的取值时传播时间 n 的多项式的下界。

3.6 本章小结

本章介绍了理论分析之前在 Kleinberg 小世界网络上做的蒙特卡洛实验以及结果分析。3.1 节介绍了实验的设计, 简要的列出了模拟时采用的算法。3.2 节介绍了服务器的配置, 3.3 节和 3.4 节介绍了实验目的到实验步骤, 包括参数的设置和网络大小的选取。3.5 节用图表的形式展示了实验结果, 并分析了实验图像, 探究了传播时间和网络大小 n 以及模型参数 α 的关系。



4 复杂传播的研究

本章在第三章蒙特卡洛模拟实验的基础上, 给出复杂传播完整的理论上的分析。首先给出我们关于复杂传播的主要结论, 然后根据不同的 α 的范围, 分别证明。

4.1 复杂传播的主要结论

本节研究了无向 Kleinberg 小世界网络 (与^[15]保持一致, 弱连接是无向的) 中的复杂传播, 并给出了期望传播时间的下界以及高概率下界 (例如以很高的概率 $1 - O(n^{-\varepsilon})$ 传播时间大于这个下界)。所有在无向 Kleinberg 小世界网络中成立的下界在有向 Kleinberg 小世界网络中依然成立, 因为在弱连接为有向边的网络中传播速度会比在弱连接无向的网络中慢。在本文中, 主要定理和分析都是基于 2-复杂传染病模型, 探讨的 Kleinberg 的小世界是基于二维网格的网络。在最后, 本章会列出对于 k -复杂传染病模型的通用结论。

定理 4.1: 对于任意的比较小的 $\varepsilon > 0$, 在无向 Kleinberg 小世界网络中, 2-复杂传染病的传播时间具有依赖于 Kleinberg 小世界模型参数 α 的下界:

- (1) 对于 $\alpha \in [0, 1/5)$, 传播时间是 $\Omega(n^{\frac{1-\alpha-\varepsilon}{4}})$ 的概率至少为 $1 - O(n^{-\varepsilon})$, 期望传播时间的下界为 $\Omega(n^{\frac{1-\alpha}{4}})$ 。
- (2) 对于 $\alpha \in [1/5, 4/5]$, 传播时间是 $\Omega(n^{\frac{1-\varepsilon}{5}})$ 的概率至少为 $1 - O(n^{-\varepsilon})$, 期望传播时间的下界为 $\Omega(n^{\frac{1}{5}})$ 。
- (3) 对于 $\alpha \in (4/5, 2)$, 传播时间是 $\Omega(n^{\frac{2-\alpha-\varepsilon}{6}})$ 的概率至少为 $1 - O(n^{-\varepsilon})$, 期望传播时间的下界为 $\Omega(n^{\frac{2-\alpha}{6}})$ 。
- (4) 对于 $\alpha \in (3, +\infty)$, 传播时间是 $\Omega(n^{\frac{\alpha-3-\varepsilon}{2\alpha}})$ 的概率至少为 $1 - O(n^{-\varepsilon})$, 期望传播时间的下界为 $\Omega(n^{\frac{\alpha-3}{2\alpha}})$ 。

和在第 2 章介绍的一致, 为了研究 2-复杂传染病模型, 本文取网络上曼哈顿距离为 1 的两个节点作为种子节点, 用 $S_0 = \{s_0^1, s_0^2\}$ 来表示这两个种子节点。对于 Kleinberg 小世界模型, 每个节点 u 和曼哈顿距离 2 以内的所有节点建立强连接 ($p = 2$), u 也会发出两条弱连接 ($q = 2$)。注意到每个节点发出两条弱连接, 但是实际上可能与许多条弱连接相连, 因为节点可以收到来自其他节点发出的弱连接。



4.2 分析与证明

用集合序列 $S_0, S_1, \dots, S_\ell = V$ 表示在复杂传播过程中被感染的节点集合序列, 其中 S_i 表示在时间步骤 i 时当前的被感染的节点集合, ℓ 指所有节点即 V 被感染时的时间, 也就是复杂传播需要的时间。因此, S_i 的大小是递增的, 而且有 $S_0 \subseteq S_1 \subseteq \dots \subseteq S_\ell$ 。用 B_i 表示以节点 s_0^1 为圆心, 半径为 $c(i+1)n^\delta$ 的“圆”中所有的节点 (这个圆是在二维网格上画出的)。 B_i 的数学表达为 $B_i = \{x \in V \mid |s_0^1 x| \leq (i+1) \cdot c \cdot n^\delta\}$, 其中 $i = 0, 1, \dots$, δ 是后文中将要确定的非负参数, c 是用来保证 $c \cdot n^\delta > 2$ 的常数。这样, 两个相邻的圆 (B_i, B_{i+1}) 半径至少相差 2。因为 $|s_0^1 s_0^2| = 1$, 所以 $S_0 \subseteq B_0$ 是显然成立的。

在这里定义一些符号的含义:

- 1、 $u \xleftrightarrow{\geq 2} S$: 节点 u 与集合 S 中的 2 个节点分别发出的两条弱连接相连。
- 2、 $u \xleftrightarrow{\geq 1} v$: 节点 u 和节点 v 之间存在一条弱连接, 这条弱连接发自 u 或者 v 。
- 3、 $u \rightarrow v$: 节点 u 发出了一条终点为节点 v 的弱连接。

设 $p(u, v)$ 为对于 u 发出的一条弱连接, 终点为 v 的概率。对于不同的模型参数 α , 归一化因子 \mathcal{Z} 的值不同, 下面列出归一化因子的值以及对应的概率 $p(u, v)$ 。

- 1、 $\mathcal{Z} = \Theta(\frac{1}{n^{1-\alpha/2}})$, $p(u, v) = \Theta(\frac{1}{n^{1-\alpha/2}|uv|^\alpha})$, $\alpha \in [0, 2)$;
- 2、 $\mathcal{Z} = \Theta(\frac{1}{\log n})$, $p(u, v) = \Theta(\frac{1}{|uv|^2 \log n})$, $\alpha = 2$;
- 3、 $\mathcal{Z} = \Theta(1)$, $p(u, v) = \Theta(\frac{1}{|uv|^\alpha})$, $\alpha \in (2, +\infty)$ 。

4.2.1 定理4.1 (1) 的证明

为了证明 $0 \leq \alpha < 1/5$ 时复杂传播时间的下界, 首先需要下面这条引理。

引理 4.1: 在无向 *Kleinberg* 小世界网络中选定三个节点 x, y, z (三个节点可以相同), 节点 z 同时与 x 和 y 通过边相连的概率 $\Pr(x \xleftrightarrow{\geq 1} z, y \xleftrightarrow{\geq 1} z)$ 不超过 $14p(x, z)p(y, z)$ 。

证明 最开始假设节点 x, y, z 都是不同的。为了计算概率 $\Pr(x \xleftrightarrow{\geq 1} z, y \xleftrightarrow{\geq 1} z)$, 首先需要把事件 $\{x \xleftrightarrow{\geq 1} z, y \xleftrightarrow{\geq 1} z\}$ 拆分成四个较小的事件的并。通过布尔不等式 (*Boole's*



inequality, 也称作 *Union bound*) 可以得到

$$\begin{aligned}
 & \Pr(x \overset{\geq 1}{\longleftrightarrow} z, y \overset{\geq 1}{\longleftrightarrow} z) \\
 & \leq \Pr(x \overset{\geq 1}{\longleftrightarrow} z, y \overset{\geq 1}{\longleftrightarrow} z, x \rightarrow z, y \rightarrow z) + \Pr(x \overset{\geq 1}{\longleftrightarrow} z, y \overset{\geq 1}{\longleftrightarrow} z, x \rightarrow z, z \rightarrow y) \\
 & \quad + \Pr(x \overset{\geq 1}{\longleftrightarrow} z, y \overset{\geq 1}{\longleftrightarrow} z, z \rightarrow x, y \rightarrow z) + \Pr(x \overset{\geq 1}{\longleftrightarrow} z, y \overset{\geq 1}{\longleftrightarrow} z, z \rightarrow x, z \rightarrow y) \\
 & \leq \Pr(x \rightarrow z, y \rightarrow z) + \Pr(x \rightarrow z, z \rightarrow y) + \Pr(z \rightarrow x, y \rightarrow z) + \Pr(z \rightarrow x, z \rightarrow y).
 \end{aligned}$$

因为节点 x, y, z 是固定的节点, 他们彼此之间发出的弱连接是互相独立的。于是有如下概率的等式 $\Pr(x \rightarrow z, y \rightarrow z) = \Pr(x \rightarrow z, z \rightarrow y) = \Pr(z \rightarrow x, y \rightarrow z)$ (但是 $\Pr(z \rightarrow x, z \rightarrow y)$ 和他们不同, 因为两条弱连接都是从 z 发出的)。下面以计算 $\Pr(x \rightarrow z, y \rightarrow z)$ 的值作为代表, $\Pr(x \rightarrow z, z \rightarrow y), \Pr(z \rightarrow x, y \rightarrow z)$ 的计算方法和结果都是一样的, 此处就不再列出。

节点 x 发出两条弱连接, 每一条都会以 $p(x, z)$ 的概率指向节点 z , 所以 $\Pr(x \rightarrow z) \leq 2p(x, z)$ 。因为节点 x 发出的弱连接与节点 y 发出的弱连接相互独立, 所以事件 $\{x \rightarrow z\}$ 和事件 $\{y \rightarrow z\}$ 也是独立的。因此 $\Pr(x \rightarrow z, y \rightarrow z) = \Pr(x \rightarrow z) \Pr(y \rightarrow z) \leq 4p(x, z)p(y, z)$ 。

接下来讨论概率 $2\Pr(z \rightarrow x, z \rightarrow y)$, 首先可以把事件 $\{z \rightarrow x, z \rightarrow y\}$ 拆分成两个事件。第一个事件是 z 用第一条弱连接指向 x , 用第二条弱连接指向 y 。第二个事件是 z 用第二条弱连接指向 x , 用第一条弱连接指向 y 。因为事件 $\{z \rightarrow x\}$ 和事件 $\{z \rightarrow y\}$ 是独立的, 利用独立事件的概率公式可以得到

$$\Pr(z \rightarrow x, z \rightarrow y) \leq p(z, x)p(z, y) + p(z, y)p(z, x) = 2p(z, x)p(z, y).$$

很显然 $p(x, z) = p(z, x)$ 并且 $p(y, z) = p(z, y)$ 。所以通过以上分析得到了一个节点 z 被其他节点通过两条弱连接相连的概率的不等式

$$\Pr(x \overset{\geq 1}{\longleftrightarrow} z, y \overset{\geq 1}{\longleftrightarrow} z) \leq 14p(x, z)p(y, z).$$

注意到如果 $x = z$ 或者 $y = z$, $\Pr(x \overset{\geq 1}{\longleftrightarrow} z, y \overset{\geq 1}{\longleftrightarrow} z) = 0$, 这是因为弱连接不会指向自己。对于 $x = y$ 的情况, $\{x \overset{\geq 1}{\longleftrightarrow} z, y \overset{\geq 1}{\longleftrightarrow} z\}$ 意味着节点 x 和节点 z 之间至少有两弱连接 $\{x \overset{\geq 2}{\longleftrightarrow} z\}$ 。节点 x 和节点 z 各发出 2 条弱连接, 于是 $\Pr(x \overset{\geq 2}{\longleftrightarrow} z) \leq \binom{4}{2}p(x, z)p(x, z)$,



引理中的不等式依然成立 □

在引理4.1的基础上，下面本文给出定理4.1(1) 的证明。

证明 (定理4.1 (1)) 首先证明， V 中存在被与 B_0 的两条弱连接（弱连接可以来自 B_0 ，也可以来自 z ）相连的点 z 的概率很低，即 $\Pr(\exists z \in V, z \xleftrightarrow{\geq 2} B_0)$ 比较小。

$$\begin{aligned}
 & \Pr(\exists z \in V, z \xleftrightarrow{\geq 2} B_0) \\
 &= \Pr(\exists z \in V, \exists x, y \in B_0, x \neq y, x \xleftrightarrow{\geq 1} z, y \xleftrightarrow{\geq 1} z) \\
 &\leq \sum_{z \in V} \sum_{x, y \in B_0} \Pr(x \xleftrightarrow{\geq 1} z, y \xleftrightarrow{\geq 1} z) \\
 &\leq |V| \cdot |B_0|^2 \Pr(x \xleftrightarrow{\geq 1} z, y \xleftrightarrow{\geq 1} z) \\
 &\leq 16c^4 n^{1+4\delta} \Pr(x \xleftrightarrow{\geq 1} z, y \xleftrightarrow{\geq 1} z).
 \end{aligned} \tag{4.1}$$

最后一步的不等式是因为 $|V| = n$ 而且 $|B_0| \leq (2cn^\delta)^2 = 4c^2 n^{2\delta}$ 。

从引理4.1，可以得到 $\Pr(x \xleftrightarrow{\geq 1} z, y \xleftrightarrow{\geq 1} z) \leq 14p(x, z)p(y, z)$ 。因为 $\alpha < 1/5$ ，归一化因子为 $\mathcal{Z} = \Theta(\frac{1}{n^{1-\alpha/2}})$ 。两个节点之间的曼哈顿距离一定有 $|xz| \geq 1, |yz| \geq 1$ ，所以带入上面的不等式中可以得到

$$\Pr(x \xleftrightarrow{\geq 1} z, y \xleftrightarrow{\geq 1} z) \leq 14\mathcal{Z}^2 = O(\frac{1}{n^{2-\alpha}}).$$

在此处可以设置 $\delta = \frac{1-\alpha-\epsilon}{4}$ ，然后得到了以下的结果：

$$\Pr(\exists z \in V, z \xleftrightarrow{\geq 2} B_0) \leq 16c^4 n^{1+4\delta} \Pr(x \xleftrightarrow{\geq 1} z, y \xleftrightarrow{\geq 1} z) = 16c^4 n^{1+4\delta} O(\frac{1}{n^{2-\alpha}}) = O(n^{-\epsilon}).$$

因为 $\Pr(\exists z \in V, z \xleftrightarrow{\geq 2} B_0)$ 很小，于是可以说对于任意的 *Kleinberg* 小世界网络，当模型参数 $\alpha < 1/5$ 时，网络中存在与 B_0 通过两条弱连接相连的概率很小。在这个条件下， B_0 外有节点被感染之前复杂传播都很慢。这是因为节点被感染只能通过（1）仅靠强连接，（2）强连接和弱连接的结合。不管怎样都需要一条强连接的帮助才能感染新的节点，而强连接的长度不超过 2。所以在时间 i ，被感染的节点集合 S_i 最多只能感染那些距离 S_i 曼哈顿距离不超过 2 的节点。因此有很高的概率在复杂传播的开始阶段，被感染的节点集合在每一步只能向外扩展 2 个曼哈顿距离。更正式地说，在 S_i 的半径小于 cn^δ 时， S_i 每一步最多把半径增加 2 的概率至少为 $1 - O(n^{-\epsilon})$ 。而仅仅感染集合 B_0

需要的时间就至少为 $\frac{cn^\delta}{2} = \Omega(n^{\frac{1-\alpha-\varepsilon}{4}})$, 因此复杂传播的时间以 $1 - O(n^{-\varepsilon})$ 的概率至少为 $1 - O(n^{-\varepsilon})$ 。

对于期望传播时间的下界, 需要更精确的计算才能得到。由于在 $\alpha < 2$ 时, 模型的归一化因子为 $\mathcal{Z} = \Theta(n^{1-\alpha/2})$ 。因此存在一个常数 β , 有 $\mathcal{Z} \leq \beta \frac{1}{n^{1-\alpha/2}}$ 。令 $c = \frac{1}{4\sqrt{\beta}}, \delta = \frac{1-\alpha}{4}$, 在 n 足够大时仍然有 $cn^\delta > 2$, 因此 c 和 δ 的取值是合适的。下面把 c, δ, \mathcal{Z} 的值带入不等式 (4.1) 中,

$$\Pr(\exists z \in V, z \xleftrightarrow{\geq 2} B_0) \leq 16c^4 n^{1+4\delta} \Pr(x \xleftrightarrow{\geq 1} z, y \xleftrightarrow{\geq 1} z) \leq 16c^4 n^{1+4\delta} \frac{14\beta^2}{n^{2-\alpha}} \leq \frac{7}{8}.$$

从上面关于概率的不等式可以得到: 以至少 $\frac{1}{8}$ 的概率, 传播时间至少为 $\frac{cn^{\frac{1-\alpha}{4}}}{2}$ 。那么 2-复杂传播的期望传播时间为 $\Omega(\frac{1}{8} \cdot \frac{cn^{\frac{1-\alpha}{4}}}{2} + \frac{7}{8} \cdot 1) = \Omega(n^{\frac{1-\alpha}{4}})$ 。□

这里 $\alpha = 0$ 复杂传播时间的下界和 Ghasemiesfeh 等人给出的 Newman-Watts 模型中复杂传播时间的下界相符合^{[15] 1}。

4.2.2 证明方法框架

前面一节采用的证明方法对于比较大的模型参数 α , 例如 $\alpha \geq 1$ 就不合适了。这一小节提供了一个比较通用的证明方法来更好的研究复杂传播, 同时在 $\alpha \geq 1/5$ 时也得到了比较高的下界。主要思路是用一个一个的“圆”来覆盖住已感染的节点集合 S_i 。最终要证明, 以很高的概率, 在最初的 $i < \lambda n^\gamma$ 步中, $S_i \subseteq B_i$ 一直成立。此处 $\lambda < 1$ 和 $\gamma \leq \frac{1}{2}$ 是接下来要确定的常数。只要保证 $\lambda n^\gamma \times cn^\delta \leq \sqrt{n}$, 圆 B_i 的个数 $\Omega(n^\gamma)$ 就可以作为 2-复杂传播时间的下界。

定义事件 $\xi = \{\forall 1 \leq i < \lambda n^\gamma, S_i \subseteq B_i\}$, 接下来要证明事件 ξ 发生的概率很低, 或者等价地说概率 $1 - \Pr(\xi)$ 非常小。这也就是说, 在传播开始的 λn^γ 步中, 每一步已感染的节点集最多对向外扩张 cn^δ 曼哈顿距离。

定义 $R_l(s_0^1)$ 为距离节点 s_0^1 曼哈顿距离为 l 的节点集合, 这就像是以 s_0^1 为圆心 l 为半径的圆环的上的节点。不难证明 $l \leq |R_l(s_0^1)| \leq 4l$ 。

现在本文给出事件 ξ 概率的不等式。假设在第 i 步, 有 $S_{i-1} \subseteq B_{i-1}$ 。如果节点 $z \in V - B_i$ 想要被感染, z 一定要通过两条弱连接和 S_{i-1} 相连。这是因为 B_i 间的距离 $c \cdot n^\delta > 2$, 不通过弱连接去感染的话, S_{i-1} 只能感染距离 S_{i-1} 不超过 2 的节点。因此,

¹Ghasemiesfeh 等人提供的 Newman-Watts 模型中复杂传播的下界 $\Omega(n^{\frac{1}{3}})$ ^[15] 是在一维的圆环模型上。扩展到二维网格模型中, 下界实际上是 $\Omega(n^{\frac{1}{6}})$ 。用他们的证明方法^[15], 实际上可以得到在二维网格模型中, 期望的传播时间为 $\Omega(n^{\frac{1}{4}})$, 这我们的结果符合。



当节点 $z \notin B_i$ 被 $z \notin B_i$ 感染时, 可以推断出 $z \xrightarrow{\geq 2} S_{i-1}$ 。因此

$$\begin{aligned}
 & 1 - \Pr(\xi) \\
 &= \Pr(\exists 1 \leq i < \lambda n^\gamma, S_i \not\subseteq B_i) \\
 &\leq \sum_{i=1}^{\lambda n^\gamma-1} \Pr(S_{i-1} \subseteq B_{i-1}, S_i \not\subseteq B_i) \\
 &= \sum_{i=1}^{\lambda n^\gamma-1} \Pr(\exists z \notin B_i, S_{i-1} \subseteq B_{i-1}, z \in S_i) \\
 &= \sum_{i=1}^{\lambda n^\gamma-1} \Pr(\exists z \notin B_i, S_{i-1} \subseteq B_{i-1}, z \xrightarrow{\geq 2} S_{i-1}) \\
 &\leq \sum_{i=1}^{\lambda n^\gamma-1} \Pr(\exists z \notin B_i, z \xrightarrow{\geq 2} B_{i-1}) \\
 &= \sum_{i=1}^{\lambda n^\gamma-1} \Pr(\exists z \notin B_i, \exists x, y \in B_{i-1}, x \neq y, x \xrightarrow{\geq 1} z, y \xrightarrow{\geq 1} z) \\
 &\leq \sum_{i=1}^{\lambda n^\gamma-1} \sum_{x, y \in B_{i-1}} \sum_{z \notin B_i} \Pr(x \xrightarrow{\geq 1} z, y \xrightarrow{\geq 1} z).
 \end{aligned} \tag{4.2}$$

和上一节的证明方法不同, 本节不再一视同仁的对待 B_i 中的节点 z 。因为 z 与 S_{i-1} 的距离并不相同。在本节的证明框架中, 把集合 $V - B_i$ 分成了一些列的不同半径的圆环, 然后考虑 z 在不同的圆环上与 S_{i-1} 的距离。这样可以得到有关 $\Pr(x \xrightarrow{\geq 1} z, y \xrightarrow{\geq 1} z)$ 更精确的不等式。

$$\sum_{z \notin B_i} \Pr(x \xrightarrow{\geq 1} z, y \xrightarrow{\geq 1} z) = \sum_{l=(i+1) \cdot cn^\delta}^{\sqrt{n}} \sum_{z \in R_l(s_0^1)} \Pr(x \xrightarrow{\geq 1} z, y \xrightarrow{\geq 1} z)$$

由距离的三角不等式可以得到 $|xz| \geq l - i \cdot cn^\delta$ 和 $|yz| \geq l - i \cdot cn^\delta$ 。把有关 $|xz|, |yz|$ 曼哈顿距离的不等式带入引理4.1得到了更精确的不等式

$$\Pr(x \xrightarrow{\geq 1} z, y \xrightarrow{\geq 1} z) \leq 14p(x, z)p(y, z) = \frac{14\mathcal{Z}^2}{|xz|^\alpha |yz|^\alpha} \leq 14\mathcal{Z}^2(l - icn^\delta)^{-2\alpha}.$$



因此可以给出 $\sum_{z \notin B_i} \Pr(x \xrightarrow{\geq 1} z, y \xrightarrow{\geq 1} z)$ 的不等式

$$\begin{aligned}
 & \sum_{l=(i+1) \cdot cn^\delta}^{\sqrt{n}} \sum_{z \in R_l(s_0^1)} \Pr(x \xrightarrow{\geq 1} z, y \xrightarrow{\geq 1} z) \\
 & \leq \sum_{l=(i+1) \cdot cn^\delta}^{\sqrt{n}} |R_l(s_0^1)| \cdot 14\mathcal{Z}^2(l - icn^\delta)^{-2\alpha} \\
 & \leq 14\mathcal{Z}^2 \sum_{l=(i+1) \cdot cn^\delta}^{\sqrt{n}} 4l \cdot (l - icn^\delta)^{-2\alpha} \\
 & = 56\mathcal{Z}^2 \sum_{l=cn^\delta}^{\sqrt{n}-icn^\delta} (l + icn^\delta) \cdot l^{-2\alpha} \\
 & = 56\mathcal{Z}^2 \left(\sum_{l=cn^\delta}^{\sqrt{n}-icn^\delta} l^{1-2\alpha} + icn^\delta \sum_{l=cn^\delta}^{\sqrt{n}-icn^\delta} l^{-2\alpha} \right).
 \end{aligned}$$

结合上面的不等式和不等式4.2, 得到对于事件 ξ 的概率:

$$\begin{aligned}
 & 1 - \Pr(\xi) \\
 & \leq \sum_{i=1}^{\lambda n^\gamma - 1} \sum_{x, y \in B_{i-1}} \sum_{l=(i+1) \cdot cn^\delta}^{\sqrt{n}} \sum_{z \in R_l(s_0^1)} \Pr(x \xrightarrow{\geq 1} z, y \xrightarrow{\geq 1} z) \\
 & \leq \sum_{i=1}^{\lambda n^\gamma - 1} |B_{i-1}|^2 \cdot 56\mathcal{Z}^2 \left(\sum_{l=cn^\delta}^{\sqrt{n}-icn^\delta} l^{1-2\alpha} + icn^\delta \sum_{l=cn^\delta}^{\sqrt{n}-icn^\delta} l^{-2\alpha} \right) \quad (4.3) \\
 & \leq \lambda n^\gamma \cdot 16c^4 \lambda^4 n^{4\gamma+4\delta} \cdot 56\mathcal{Z}^2 \left(\sum_{l=cn^\delta}^{\sqrt{n}-1} l^{1-2\alpha} + c\lambda n^\gamma \cdot n^\delta \sum_{l=cn^\delta}^{\sqrt{n}-1} l^{-2\alpha} \right) \\
 & = 7 \cdot 2^7 c^4 \lambda^5 n^{5\gamma+4\delta} \mathcal{Z}^2 \left(\sum_{l=cn^\delta}^{\sqrt{n}-1} l^{1-2\alpha} + c\lambda n^{\gamma+\delta} \sum_{l=cn^\delta}^{\sqrt{n}-1} l^{-2\alpha} \right).
 \end{aligned}$$

从第三行到第四行是因为 $|B_{i-1}| \leq (2cin^\delta)^2 = 4c^2 i^2 n^{2\delta}$ 并且 $i < \lambda n^\gamma$.

4.2.3 定理4.1 (2),(3),(4) 的证明

本小节提供了定理4.1余下部分的证明。证明的过程基于两个级数（见下面）的累加和。根据不同 α 对应的级数不同的累加和，本文分别给出 $\alpha \in [1/5, 1/2)$, $\alpha = 1/2$,

$\alpha \in (1/2, 4/5]$, $\alpha \in (4/5, 1)$, $\alpha = 1$, $\alpha \in (1, 2)$ 和 $\alpha \in (3, \infty)$ 的复杂传播时间的下界。

$$\sum_{l=cn^\delta}^{\sqrt{n}-1} l \cdot l^{-2\alpha} \leq \int_{cn^\delta-1}^{\sqrt{n}} x^{1-2\alpha} dx = \begin{cases} O(n^{1-\alpha}), & \alpha \in [0, 1); \\ O(\log n), & \alpha = 1; \\ O(n^{\delta(2-2\alpha)}), & \alpha \in (1, \infty). \end{cases} \quad (4.4)$$

$$\sum_{l=cn^\delta}^{\sqrt{n}-1} l^{-2\alpha} \leq \int_{cn^\delta-1}^{\sqrt{n}} x^{-2\alpha} dx = \begin{cases} O(n^{\frac{1}{2}-\alpha}), & \alpha \in [0, 1/2); \\ O(\log n), & \alpha = 1/2; \\ O(n^{\delta(1-2\alpha)}), & \alpha \in (1/2, \infty). \end{cases} \quad (4.5)$$

证明 (定理4.1 (2),(3),(4)) 当 $\alpha \in [1/5, 1/2)$ 时, 归一化因子为 $\mathcal{Z} = \Theta(\frac{1}{n^{1-\alpha/2}})$ 。令 $\delta = 0$, 也就是说相邻的 B_i 之间半径相差为大于 2 的常数。把级数的和带入不等式 (4.3), 得到

$$\begin{aligned} 1 - \Pr(\xi) &\leq 7 \cdot 2^7 c^4 \lambda^5 n^{5\gamma+4\delta} \mathcal{Z}^2 \left(\sum_{l=cn^\delta}^{\sqrt{n}-1} l^{1-2\alpha} + c\lambda n^{\gamma+\delta} \sum_{l=cn^\delta}^{\sqrt{n}-1} l^{-2\alpha} \right) \\ &\leq O(n^{5\gamma}) \cdot O(\frac{1}{n^{2-\alpha}}) (O(n^{1-\alpha}) + \sqrt{n} O(n^{1/2-\alpha})) = O(\frac{n^{5\gamma}}{n}). \end{aligned}$$

这里可取 $\gamma = \frac{1-\varepsilon}{5}$ 并且 $\lambda = \frac{1}{c}$, ε 是一个比较小的正数。不难发现 γ 和 λ 的取值满足 $\lambda n^\gamma \times cn^\delta < \sqrt{n}$ 。在确定了 γ 和 λ 的值后, $1 - \Pr(\xi) \leq O(n^{-\varepsilon})$, 这样就意味着有很高的概率 $1 - O(n^{-\varepsilon})$, 在复杂传播最初 $\lambda n^\gamma - 1$ 步中每一步已感染节点集合 S_i 的半径都小于 $(i+1) \cdot cn^\delta$ 。由于 $c\lambda n^\gamma n^\delta < \sqrt{n}$, 被感染的节点的集合在第 $\lambda n^\gamma - 1$ 步时仍然是全部节点结合 V 的子集。因此, 以很高的概率 $1 - O(n^{-\varepsilon})$, 2-复杂传播在第 $\lambda n^\gamma = \frac{1}{c} n^{\frac{1-\varepsilon}{5}}$ 步时仍然继续, 所以高概率下界成立了。

对于复杂传播时间的期望, 需要进一步对这些常数进行赋值。

$$\begin{aligned} \sum_{l=cn^\delta}^{\sqrt{n}-1} l^{1-2\alpha} &\leq \int_{cn^\delta-1}^{\sqrt{n}} x^{1-2\alpha} dx \leq \frac{(\sqrt{n})^{2-2\alpha}}{2-2\alpha}, \\ \sum_{l=cn^\delta}^{\sqrt{n}-1} l^{-2\alpha} &\leq \int_{cn^\delta-1}^{\sqrt{n}} x^{-2\alpha} dx < \frac{(\sqrt{n})^{1-2\alpha}}{1-2\alpha}. \end{aligned}$$

注意到对于归一化因子, 一定存在 $\beta > 1$ 使得 $\mathcal{Z} \leq \beta \frac{1}{n^{1-\alpha/2}}$ 。把级数的和以及归一化



因子的不等式代入 $\Pr(\xi)$, 得到

$$\begin{aligned}
 & 1 - \Pr(\xi) \\
 & \leq 7 \cdot 2^7 c^4 \lambda^5 n^{5\gamma+4\delta} \mathcal{Z}^2 \left(\sum_{l=cn^\delta}^{\sqrt{n}-1} l^{1-2\alpha} + c\lambda n^{\gamma+\delta} \sum_{l=cn^\delta}^{\sqrt{n}-1} l^{-2\alpha} \right) \\
 & \leq 7 \cdot 2^7 c^4 \lambda^5 n^{5\gamma+4\delta} \cdot \beta^2 \frac{1}{n^{2-\alpha}} \left(\frac{n^{1-\alpha}}{2-2\alpha} + \sqrt{n} \frac{n^{\frac{1}{2}-\alpha}}{1-2\alpha} \right) \\
 & \leq \frac{7 \cdot 2^7 c^4 \lambda^5 n^{5\gamma+4\delta} \beta^2}{n^{2-\alpha}} \cdot 2 \frac{n^{1-\alpha}}{1-2\alpha} \\
 & = \frac{7 \cdot 2^8 c^4 \lambda^5 \beta^2}{1-2\alpha} \cdot \frac{n^{5\gamma}}{n}.
 \end{aligned}$$

令 $\lambda = \frac{(1-2\alpha)^{\frac{1}{5}}}{2^{\frac{11}{5}} c \beta^{\frac{2}{5}}}$, 同时 γ 的取值为 $\gamma = \frac{1}{5}$. 很显然 $\lambda n^\gamma \times cn^\delta < \sqrt{n}$ 是成立的。因此本文得到 $1 - \Pr(\xi) \leq \frac{7}{8c}$, 这是比较小的常数。因此以 $1 - \frac{7}{8c}$ 的概率, 传播时间大于 $\frac{(1-2\alpha)^{\frac{1}{5}}}{2^{\frac{11}{5}} c \beta^{\frac{2}{5}}} n^{\frac{1}{5}}$ 。这也就是说, 复杂传播时间的期望为 $\Omega(n^{\frac{1}{5}})$ 。这样就得到了对于 $\alpha \in [1/5, 1/2)$ 的 *Kleinberg* 小世界网路中复杂传播时间期望的下界。

对于其他 α 的范围, $\alpha \in [1/2, 2)$ 和 $\alpha > 3$, 也可以分别选择合适的 λ 来得到事件 ξ 发生的概率的常数值, 然后就可以给出期望传播时间的下界。因此本小节在接下来的部分略过了期望传播时间的证明, 而仅提供高概率下界的证明。

1、当 $\alpha = 1/2$ 时, 令 $\delta = 0$, 然后取 γ 和 λ 为 $\gamma = \frac{1-\varepsilon}{5}$ 并且 $\lambda = \frac{1}{c}$, 因此得到了

$$\begin{aligned}
 & 1 - \Pr(\xi) \\
 & \leq 7 \cdot 2^7 c^4 \lambda^5 n^{5\gamma+4\delta} \mathcal{Z}^2 \left(\sum_{l=cn^\delta}^{\sqrt{n}-1} l^{1-2\alpha} + c\lambda n^{\gamma+\delta} \sum_{l=cn^\delta}^{\sqrt{n}-1} l^{-2\alpha} \right) \\
 & \leq O(n^{5\gamma}) \cdot O\left(\frac{1}{n^{3/2}}\right) \left(O(n^{1/2}) + O(n^\gamma) \cdot O(\log n) \right) \\
 & = O\left(\frac{n^{5\gamma}}{n}\right) + O\left(\frac{n^{6\gamma} \log n}{n^{3/2}}\right) \\
 & = O(n^{-\varepsilon}).
 \end{aligned}$$

事件 ξ 发生的概率至少为 $1 - O(n^{-\varepsilon})$, 因此可以得知对于 $\alpha = 1/2$ 复杂传播时间高概率下界为 $\Omega(n^{\frac{1-\varepsilon}{5}})$



2、 当 $\alpha \in (1/2, 1)$ 时, 令 $\delta = 0$, $\lambda = \frac{1}{c}$ 。不等式 (4.3) 变成:

$$\begin{aligned}
 & 1 - \Pr(\xi) \\
 & \leq 7 \cdot 2^7 c^4 \lambda^5 n^{5\gamma+4\delta} \mathcal{Z}^2 \left(\sum_{l=cn^\delta}^{\sqrt{n}-1} l^{1-2\alpha} + c\lambda n^{\gamma+\delta} \sum_{l=cn^\delta}^{\sqrt{n}-1} l^{-2\alpha} \right) \\
 & \leq O(n^{5\gamma}) \cdot O\left(\frac{1}{n^{2-\alpha}}\right) \left(O(n^{1-\alpha}) + O(n^\gamma) O(n^{\delta(1-2\alpha)}) \right) \\
 & \leq O(n^{5\gamma}) \cdot O\left(\frac{1}{n^{2-\alpha}}\right) \left(O(n^{1-\alpha}) + c\lambda n^\gamma O(1) \right) \\
 & \leq O\left(\frac{n^{5\gamma}}{n} + \frac{n^{6\gamma}}{n^{2-\alpha}}\right).
 \end{aligned}$$

对于 $\alpha \in (1/2, 4/5]$, 本文取 $\gamma = \frac{1-\varepsilon}{5}$. 因此 $1 - \Pr(\xi) \leq O\left(\frac{1}{n^\varepsilon}\right) + O\left(\frac{1}{n^{\frac{4}{5}-\alpha+\frac{6\varepsilon}{5}}}\right) = O(n^{-\varepsilon})$.
 对 $\alpha \in (1/2, 4/5]$ 的 *Kleinberg* 小世界网络以 $1 - O(n^{-\varepsilon})$ 的概率传播时间为 $\Omega(n^{\frac{1-\varepsilon}{5}})$.
 至于 $\alpha \in (4/5, 1)$ 时, 取 $\gamma = \frac{2-\alpha-\varepsilon}{6}$. 可以得到事件 ξ 的概率 $1 - \Pr(\xi) = O\left(\frac{1}{n^{1-\frac{5(2-\alpha)}{6}+\frac{5\varepsilon}{6}}}\right) + O\left(\frac{1}{n^\varepsilon}\right) = O(n^{-\varepsilon})$. 此时传播时间高概率下界为 $\Omega(n^{\frac{2-\alpha-\varepsilon}{6}})$.

3、 当 $\alpha = 1$ 时, 令 $\delta = 0$, $\gamma = \frac{1-\varepsilon}{6}$, $\lambda = \frac{1}{c}$ 。然后计算事件 ξ 发生的概率。

$$\begin{aligned}
 & 1 - \Pr(\xi) \\
 & \leq 7 \cdot 2^7 c^4 \lambda^5 n^{5\gamma+4\delta} \mathcal{Z}^2 \left(\sum_{l=cn^\delta}^{\sqrt{n}-1} l^{1-2\alpha} + c\lambda n^{\gamma+\delta} \sum_{l=cn^\delta}^{\sqrt{n}-1} l^{-2\alpha} \right) \\
 & \leq O(n^{5\gamma}) \cdot O\left(\frac{1}{n}\right) \left(O(\log n) + O(n^\gamma) \cdot O(n^{\delta(1-2\alpha)}) \right) \\
 & = O\left(\frac{n^{5\gamma} \log n}{n}\right) + O\left(\frac{n^{6\gamma}}{n}\right) \\
 & = O(n^{-\varepsilon}).
 \end{aligned}$$

于是可以得到对于 2-复杂传播的传播时间以高概率 $1 - O(n^{-\varepsilon})$ 的下界 $\Omega(n^{\frac{1-\varepsilon}{6}})$

4、 当 $\alpha \in (1, 2)$ 时, $1 - 2\alpha$ 和 -2α 都小于 -1 。令 $\delta = 0$ 然后两个级数 (4.4), (4.5) 的累

加和都为 $O(1)$ 。然后取 $\gamma = \frac{2-\alpha-\varepsilon}{6}$ 并且 $\lambda = \frac{1}{c}$,

$$\begin{aligned}
& 1 - \Pr(\xi) \\
& \leq 7 \cdot 2^7 c^4 \lambda^5 n^{5\gamma+4\delta} \mathcal{Z}^2 \left(\sum_{l=cn^\delta}^{\sqrt{n}-1} l^{1-2\alpha} + c \lambda n^{\gamma+\delta} \sum_{l=cn^\delta}^{\sqrt{n}-1} l^{-2\alpha} \right) \\
& \leq O(n^{5\gamma}) \cdot O\left(\frac{1}{n^{2-\alpha}}\right) (O(1) + n^\gamma O(1)) \\
& \leq O\left(\frac{n^{6\gamma}}{n^{2-\alpha}}\right) \\
& \leq O(n^{-\varepsilon}).
\end{aligned}$$

事件 ξ 发生的概率为 $1 - O(n^{-\varepsilon})$, 从而以很高的概率通过复杂传播感染整个网络所需要的步数至少为 $\Omega(n^{\frac{2-\alpha-\varepsilon}{6}})$ 。

- 5、当 $\alpha \in (3, +\infty)$ 时, 归一化因子 \mathcal{Z} 是一个常数。这时我们需要令 $\delta > 0$, 而不能再设置 $\delta = 0$ 。

$$\begin{aligned}
& 1 - \Pr(\xi) \\
& \leq 7 \cdot 2^7 c^4 \lambda^5 n^{5\gamma+4\delta} \mathcal{Z}^2 \left(\sum_{l=cn^\delta}^{\sqrt{n}-1} l^{1-2\alpha} + c \lambda n^{\gamma+\delta} \sum_{l=cn^\delta}^{\sqrt{n}-1} l^{-2\alpha} \right) \\
& \leq O(n^{5\gamma+4\delta}) (O(n^{\delta(2-2\alpha)}) + O(n^{\gamma+\delta}) \cdot O(n^{\delta(1-2\alpha)})) \\
& \leq O\left(\frac{n^{5\gamma+4\delta}}{n^{(2\alpha-2)\delta}} + \frac{n^{6\gamma+5\delta}}{n^{(2\alpha-1)\delta}}\right) = O\left(\frac{n^{5\gamma+6\delta} + n^{6\gamma+6\delta}}{n^{2\alpha\delta}}\right) \\
& = O\left(\frac{n^{6\gamma+6\delta}}{n^{2\alpha\delta}}\right).
\end{aligned}$$

令 $\delta = \frac{3+\varepsilon}{2\alpha}$ 并且 $\gamma = \frac{1}{2} - \delta = \frac{\alpha-3-\varepsilon}{2\alpha}$, 此时 $0 < \varepsilon < \alpha - 3$ 。如果取 $\lambda = \frac{1}{c}$, 那么 $\lambda n^\gamma \times cn^\delta \leq \sqrt{n}$ 依然成立。于是可以得到关于事件 ξ 概率的不等式 $1 - \Pr(\xi) \leq O\left(\frac{n^3}{n^{3+\varepsilon}}\right) = O(n^{-\varepsilon})$ 。因此本文可以说 2-复杂传播以 $1 - O(n^{-\varepsilon})$ 的概率中少需要 $\Omega(n^{\frac{\alpha-3-\varepsilon}{2\alpha}})$ 步才能感染整个网络。

结合关于 α 所有范围的结论, 本文得到了在 $\alpha \in [0, 2)$ 和 $\alpha \in (3, +\infty)$ 时无向 *Kleinberg* 小世界网络中复杂传播时间的高概率下界和期望传播时间的下界。 \square



4.3 讨论和扩展

对于 k -复杂传播, 本文可以得到传播时间上类似的下界。在这种情况下, 我们需要调整 Kleinberg 小世界模型的参数来保证复杂传播可以通过强连接进行下去。我们取 $p = q = k$, 同时初始的种子节点为 k 个网格上连续的节点。用上一节类似的方法我们可以得到 k -复杂传播的结果, 在这里我们只给出结论。

推论 4.1: 对于任意的比较小的 $\varepsilon > 0$, 在 $p = q = k$ 的无向 Kleinberg 小世界网络中, k -复杂传染病 ($k \geq 2$) 的传播时间具有依赖于 Kleinberg 小世界模型参数 α 的下界:

- 1、对于 $\alpha \in [0, \frac{2(k-1)}{k(2k+1)})$, 传播时间是 $\Omega(n^{\frac{2k-k\alpha-2-2\varepsilon}{4k}})$ 的概率至少为 $1 - O(n^{-\varepsilon})$, 期望传播时间的下界为 $\Omega(n^{\frac{2k-k\alpha-2}{4k}})$ 。
- 2、对于 $\alpha \in [\frac{2(k-1)}{k(2k+1)}, \frac{2k+4}{k(2k+1)}]$, 传播时间是 $\Omega(n^{\frac{k-1-\varepsilon}{2k+1}})$ 的概率至少为 $1 - O(n^{-\varepsilon})$, 期望传播时间的下界为 $\Omega(n^{\frac{k-1}{2k+1}})$ 。
- 3、对于 $\alpha \in (\frac{2k+4}{k(2k+1)}, 2)$, 传播时间是 $\Omega(n^{\frac{(2-\alpha)k-2\varepsilon}{4k+4}})$ 的概率至少为 $1 - O(n^{-\varepsilon})$, 期望传播时间的下界为 $\Omega(n^{\frac{(2-\alpha)k}{4k+4}})$ 。
- 4、对于 $\alpha \in (\frac{2k+2}{k}, +\infty)$, 传播时间是 $\Omega(n^{\frac{k\alpha-2k-2-2\varepsilon}{2k\alpha}})$ 的概率至少为 $1 - O(n^{-\varepsilon})$, 期望传播时间的下界为 $\Omega(n^{\frac{k\alpha-2k-2}{2k\alpha}})$ 。

有一点需要注意到复杂传播结果尚不明确的 α 范围是 $(2, \frac{2k+2}{k}]$, 这说明随着 k 的增加, 这个未知区间的长度会趋于 0。另外, 本文的结果对于 $k = 1$ 依然成立, $k = 1$ 的情况对应着简单传染病模型, 这个时候传播时间为网络的直径。在 $k = 1$ 时, 本文的结果在 $\alpha < 2$ 没有太大意义, 但是在 $\alpha > 4$ 时传播时间有多项式的下界, 这和小世界网络直径^[14]的结果比较类似。

4.4 本章小结

本章主要研究了 Kleinberg 小世界网络中复杂传染病的传播。首先给出了复杂传染的主要结论, 然后用列出了两种不同的证明方法, 给出了复杂传播结论的完整理论证明。最后把结论扩展到 k -复杂传播并给出了一般性的结论。

5 复杂路由的研究

本章主要研究了在有向 Kleinberg 小世界网络中复杂路由的研究, 网络模型的设定和 Kleinberg 最初提出分散式路由时的网络模型一致^[8]。正如和第二章描述的, 本章考虑节点 u 只能激活自己有向边指向的节点 (出邻居) 的分散式路由方式。只有当一个节点 u 被 k 个节点发出的有向边指向时, u 才会被变成暴露状态。对于强连接, 仍然认为它们是无向的或者双向的。在每一个时间步骤, 我们仅有知道被感染节点和被感染节点的强连接和弱连接, 而对于未感染节点的弱连接算法一无所知。这样可以在证明中使用延迟选择原则^[23], 和^[8] 在证明分散式路由中使用的原则一样。这意味着节点 u 的弱连接只有在 u 被感染时才被选取, 进而算法才知道 u 的弱连接。最初的种子节点是 k 个网格上连续的节点, 如果对于 Kleinberg 小世界网络设置 $p = q = k$, 因此 k -复杂路由最终可以感染目标节点 t 。接下来给出复杂路由的主要结论。

5.1 复杂路由的结论

我们考虑一个 2-复杂路由的任务, 起始给定一对相邻的种子节点 $S_0 = \{s_0^1, s_0^2\}$ (s_0^1 和 s_0^2 的曼哈顿距离为 1), 目标是通过分散式路由到达目标节点 t 。在本文中, 我们讨论 t 距离种子节点最远的情况, 也就是 s_0^1 与 t 的曼哈顿距离为 $\Theta(\sqrt{n})$ 。从已经暴露的节点集合中选择哪个节点来感染是由路由选择算法决定的, 一个比较特殊的路由选择策略是在已经暴露的节点中选择距离目标 t 曼哈顿距离最小的节点, 这也被称为贪心路由算法。本章的结果对于任何分散式路由选择算法都成立, 甚至是随机算法。下面是有关路由时间的下界的结果。

定理 5.1: 对于任意的分散式路由策略 (算法), 甚至是随机的策略, 对于任意的比较小的 $\varepsilon > 0$, 在有向 Kleinberg 小世界网络中, 2-复杂路由的路由时间有下面依赖于 Kleinberg 小世界模型参数 α 的下界:

- 1、当 $\alpha \in [0, 2)$ 时, 路由时间是 $\Omega(n^{\frac{1-\varepsilon}{\alpha+2}})$ 的概率至少为 $1 - O(n^{-\varepsilon})$, 期望路由时间为 $\Omega(n^{\frac{1}{\alpha+2}})$ 。
- 2、当 $\alpha = 2$ 时, 路由时间是 $\Omega(n^{\frac{1}{4}})$ 的概率至少为 $1 - O(\frac{1}{\log n})$, 期望路由时间为 $\Omega(n^{\frac{1}{4}})$ 。
- 3、当 $\alpha \in (2, +\infty)$ 时, 路由时间是 $\Omega(n^{\frac{\alpha-2\varepsilon}{2(\alpha+2)}})$ 的概率至少为 $1 - O(n^{-\varepsilon})$, 期望路由时间为 $\Omega(n^{\frac{\alpha}{2(\alpha+2)}})$ 。



在本章的开始首先给出一些定义。对于一个已感染节点集合 S ，定义集合 $E(S)$ 为被已感染节点集合 S 暴露的节点的集合，更精确的定义为 $E(S) = \{x \notin S \mid S \text{ 有两条有向边指向 } x\}$ 。从 $E(S)$ 的定义可以知道复杂路由是沿着 Kleinberg 小世界网络中的有向边扩散的。

在接下来的小节中，本文首先分析有向 Kleinberg 小世界网络中普通路由策略的性能，然后把结果推广到每一步可以感染多个节点的复杂路由。

5.2 分析与证明

首先考虑确定性的路由算法，在本节的最后，我们会证明本章的下界对于随机的路由算法依然成立。首先给出 $\alpha = 2$ 时路由时间的证明。

设 S_0, S_1, \dots, S_ℓ 为在路由过程中已感染节点集合的序列， S_i 是在时间步骤 i 时感染节点的集合。初始种子节点为 $\{s_0^1, s_0^2\}$ ，因此 $S_0 = \{s_0^1, s_0^2\}$ 。在时间步骤 $i \geq 1$ ，我们从 $E(S_{i-1})$ 中选择节点 s_i 来感染，则 S_i 和 s_i 的关系为 $S_i = \{s_0^1, s_0^2, s_1, \dots, s_i\}$ ，由于 t 最后被感染了， $s_\ell = t$ 是最后一个被加入集合 S_ℓ 的节点。定义距离 $d_i = d(S_i \cup E(S_i), t)$ ，这里 $d(S, u)$ 是指集合 S 中的节点到 u 的最小曼哈顿距离。显然 d_i 是非递增序列，而且 $d_{\ell-1} = d_\ell = 0$ 。为了方便，本文把 s_0^2 写作 s_0 ，这样可以说 s_i 是在第 i 步被感染的节点。然后定义 $d_{-1} = |s_0^1 t| = \sqrt{n}$ 因为 $|s_0^1 t| = n^{1/2}$ 。设事件 χ 为 $\chi = \{\forall 0 \leq i < cn^{1/4}, d_{i-1} - d_i \leq n^{1/4}\}$ ，此处 $c < 1$ 是一个稍后会定义的常数。 χ 是指从复杂路由的第 0 步到 $cn^{1/4} - 1$ 步，当前已感染的节点集合到目标 t 的距离每一步最多减少 $n^{1/4}$ 。接下来本文会证明当 Kleinberg 小世界模型的参数 $\alpha = 2$ 时， $\Pr(\chi)$ 是一个很高的概率，

引理 5.1: 对于 $\alpha = 2$ 的有向 Kleinberg 小世界网络，给定种子节点 $\{s_0^1, s_0^2\}$ 和最远的目标 t ($|s_0^1 t| = \sqrt{n}$)，存在一个常数 $c \in (0, 1)$ ，在分散式的 2-复杂路由的过程中有

$$\Pr(\forall 0 \leq i < cn^{1/4}, d_{i-1} - d_i \leq n^{1/4}) \geq 1 - O\left(\frac{1}{\log n}\right).$$

证明 令 u_i 为当前已感染节点集合和暴露的节点中距离目标 t 最近的节点，即 $u_i = \arg \min_x \{d(x, t) \mid x \in S_i \cup E(S_i)\}$ 。结合 d_i 的定义，可以知道 $d_i = |u_i t|$ 。因为 u_{i-1} 是集合 $E(S_{i-1}) \cup S_{i-1}$ 与目标 t 曼哈顿距离最小的节点，而 s_i 是从第 $i-1$ 步的暴露集合中选择的 $s_i \in E(S_{i-1})$ ，所以可以得到曼哈顿距离 $|s_i t|$ 的不等式 $|s_i t| \geq |u_{i-1} t| = d_{i-1}$ 。因此如果 $d_{i-1} - d_i > 0$ ，那么 u_i 必然不是 s_i 因为 s_i 到目标 t 的曼哈顿距离大于 d_{i-1} 。此外，还可以得知 $u_i \in E(S_i) - E(S_{i-1})$ (u_i 是在第 i 步刚刚变为暴露状态的节点)，同时是 s_i 在第 i 步变成了感染状态导致 u_i 变成暴露状态 (u_i 恰好在第 i 步被暴露，这正是因为 s_i 被感



染了)。结合 d_i 的定义 $|u_i t| = d_i$, 可得 $|s_i u_i| \geq |s_i t| - |u_i t| = d_{i-1} - d_i$ 。所以有了下面的结论。

如果在第 $i \geq 1$ 步有 $d_{i-1} - d_i > n^{1/4}$, 则

- 1、 $|s_i u_i| > n^{1/4}$ 。
- 2、 u_i 是 s_i 发出的有向边的终点, 实际上 s_i 发出了一条指向 u_i 的弱连接 ($|s_i u_i|$ 比较大, 两点之间只能为弱连接)。
- 3、 u_i 恰好在第 i 步变为暴露状态, 因此集合 $S_i - \{s_i\}$ 中恰好存在一条弱连接指向 u_i 。

对于 $i = 0$ 的情况因为 $d(S_0, t) = d_{-1}$, 所以是 $u_0 \in E(S_0)$ 引起了 d_{-1} 和 d_0 之间的差距, 此处的结论依然成立。

本文称那些被集合 $S_i - \{s_i\}$ 发出弱连接指向的节点集合为 X_i 。很显然根据上面的第 3 条推论 $u_i \in X_i$ 。如果在第 i 步的确有 $d_{i-1} - d_i > n^{1/4}$, 即 d_i 的值出现了突变, 那么 u_i 一定被 s_i 发出的一条长度至少为 $n^{1/4}$ 的弱连接指向。由布尔不等式, 得到

$$\begin{aligned}
 & 1 - \Pr(\forall 0 \leq i < cn^{1/4}, d_{i-1} - d_i \leq n^{1/4}) \\
 &= \Pr(\exists 0 \leq i < cn^{1/4}, d_{i-1} - d_i > n^{1/4}) \\
 &\leq \sum_{i=0}^{cn^{1/4}-1} \Pr(d_{i-1} - d_i > n^{1/4}) \\
 &\leq \sum_{i=0}^{cn^{1/4}-1} \Pr(s_i \rightarrow u_i, |s_i u_i| > n^{1/4}, u_i \in X_i) \\
 &\leq \sum_{i=0}^{cn^{1/4}-1} \Pr(\exists x \in X_i, s_i \rightarrow x, |s_i x| > n^{1/4}).
 \end{aligned} \tag{5.1}$$

因为集合 $S_i - \{s_i\}$ 中有 $i + 1$ 个节点, 所以也就发出了 $2(i + 1)$ 条弱连接。这意味着最多有 $2(i + 1)$ 个节点满足被 $S_i - \{s_i\}$ 通过弱连接指向的条件, 集合 $|X_i|$ 的大小满足 $|X_i| \leq 2(i + 1)$ 。用 $\mathcal{H}_i \subseteq 2^V$ 表示所有的节点个数不超过 $2(i + 1)$ 的结合集合, 下一步中



固定住 X_i 和 s_i 的随机性, 得到如下不等式

$$\begin{aligned}
& \Pr(\exists x \in X_i, s_i \rightarrow x, |s_i x| > n^{\frac{1}{4}}) \\
& \leq \sum_{C \in \mathcal{H}_i} \sum_{v \in V} \Pr((X_i = C) \wedge (s_i = v) \wedge (\exists x \in C, v \rightarrow x, |vx| > n^{\frac{1}{4}})) \\
& = \sum_{C \in \mathcal{H}_i} \sum_{v \in V} \Pr((X_i = C) \wedge (s_i = v)) \Pr(\exists x \in C, v \rightarrow x, |vx| > n^{\frac{1}{4}}) \\
& \leq \sum_{C \in \mathcal{H}_i} \sum_{v \in V} \Pr((X_i = C) \wedge (s_i = v)) \sum_{x \in C} \Pr(v \rightarrow x, |vx| > n^{\frac{1}{4}}) \\
& \leq \sum_{C \in \mathcal{H}_i} \sum_{v \in V} \Pr((X_i = C) \wedge (s_i = v)) \cdot |C| \cdot 2\mathcal{Z}_{\frac{1}{n^{2.1/4}}} \\
& = 2(i+1) \cdot \frac{\mathcal{Z}}{n^{1/2}} \sum_{C \in \mathcal{H}_i} \sum_{v \in V} \Pr((X_i = C) \wedge (s_i = v)) \\
& = 2(i+1) \cdot \frac{\mathcal{Z}}{n^{1/2}}.
\end{aligned}$$

由分散式路由的性质可以知道, 事件 $\{(X_i = C) \wedge (s_i = v)\}$ 仅依赖于 $S_i - \{s_i\}$ 和集合 $S_i - \{s_i\}$ 发出的弱连接。因为节点 v 不属于集合 $S_i - \{s_i\}$, 而事件 $\{\exists x \in C, v \rightarrow x, |vx| > n^{\frac{1}{4}}\}$ 又仅与固定节点 v 发出的弱连接相关。因此事件 $\{(X_i = C) \wedge (s_i = v)\}$ 和事件 $\{\exists x \in C, v \rightarrow x, |vx| > n^{\frac{1}{4}}\}$ 互相独立, 这就是公式第三行等号的来源。对于节点 x , 如果 $|vx| \leq n^{\frac{1}{4}}$, 那么概率 $\Pr(v \rightarrow x, |vx| > n^{\frac{1}{4}})$ 显然为 0; 否则的话, $\Pr(v \rightarrow x, |vx| > n^{\frac{1}{4}}) \leq 2p(v, x) \leq 2\mathcal{Z}_{\frac{1}{n^{2.1/4}}}$. 这是公式第四行到第五行 “ \leq ” 的原因。把这个不等式带入引理的不等式 (5.1) 中:

$$\begin{aligned}
& 1 - \Pr(\forall 0 \leq i < cn^{\frac{1}{4}}, d_{i-1} - d_i \leq n^{\frac{1}{4}}) \\
& \leq \sum_{i=0}^{cn^{\frac{1}{4}}-1} \Pr(\exists x \in X_i, s_i \rightarrow x, |s_i x| > n^{\frac{1}{4}}) \\
& \leq \sum_{i=0}^{cn^{\frac{1}{4}}-1} 2(i+1) \cdot \frac{\mathcal{Z}}{n^{1/2}} \\
& \leq cn^{\frac{1}{4}} \cdot 2cn^{\frac{1}{4}} \cdot \Theta\left(\frac{1}{\log n}\right) \frac{1}{n^{1/2}} \\
& = O\left(\frac{1}{\log n}\right).
\end{aligned}$$

引理得证。 □

由上面这条引理, 不难发现当 $\alpha = 2$ 时, 路由时间有很高的概率至少为 $cn^{\frac{1}{4}}$ 。下面给出本章给出的定理的证明。

证明 (定理5.1) 正如引理5.1所述, 对于 $\alpha = 2$ 的情况, 在最初的 $cn^{\frac{1}{4}}$ 步中, 已感染节点

的结合距离目标 t 在每一步最多减少 $n^{\frac{1}{4}}$ 。所以在最开始的 $cn^{\frac{1}{4}}$ 步中, 目标 t 没有被感染, 复杂路由会继续下去。以很高的概率 $1 - O(\frac{1}{\log n})$, 为了在 $\alpha = 2$ 的 *Kleinberg* 小世界网络中感染目标节点 t , 分散式的 2-复杂路由需要至少 $cn^{\frac{1}{4}}$ 步。同时期望的路由时间为 $cn^{\frac{1}{4}} \cdot (1 - O(\frac{1}{\log n})) = \Omega(n^{\frac{1}{4}})$ 。

当 $\alpha \in [0, 2)$ 时, 类比于引理 5.1 的证明方法, 可以证明对于任意的比较小的 $\varepsilon > 0$,

$$\begin{aligned}
& 1 - \Pr(\forall 0 \leq i < cn^{\frac{1-\varepsilon}{\alpha+2}}, d_{i-1} - d_i \leq 2n^{\frac{\alpha+2\varepsilon}{2(\alpha+2)}}) \\
& \leq \sum_{i=0}^{cn^{\frac{1-\varepsilon}{\alpha+2}}-1} \Pr(\exists x \in X_i, s_i \rightarrow x, |s_i x| > 2n^{\frac{\alpha+2\varepsilon}{2(\alpha+2)}}) \\
& \leq \sum_{i=0}^{cn^{\frac{1-\varepsilon}{\alpha+2}}-1} 2(i+1) \cdot \mathcal{Z} \cdot (2n^{\frac{\alpha+2\varepsilon}{2(\alpha+2)}})^{-\alpha} \\
& \leq cn^{\frac{1-\varepsilon}{\alpha+2}} \cdot 2cn^{\frac{1-\varepsilon}{\alpha+2}} \cdot \Theta(\frac{1}{n^{1-\alpha/2}}) \cdot O(n^{-\frac{\alpha(\alpha+2\varepsilon)}{2(\alpha+2)}}) \\
& = O(n^{-\varepsilon}).
\end{aligned}$$

此时复杂路由时间以 $1 - O(n^{-\varepsilon})$ 的概率至少为 $\Omega(n^{\frac{1-\varepsilon}{\alpha+2}})$ 。通过设置 $\varepsilon = 0$ 然后选取合适的常数 c , 可以得到有关路由复杂时间期望的结论, 这里不再给出证明。

当 $\alpha > 2$ 时, 我们可以像上面一样证明对于 $\alpha > 2$,

$$\Pr(\forall 0 \leq i < cn^{\frac{\alpha-2\varepsilon}{2(\alpha+2)}}, d_{i-1} - d_i \leq n^{\frac{1+\varepsilon}{\alpha+2}}) \geq 1 - O(n^{-\varepsilon}).$$

我们可以说以很高的概率 $1 - O(n^{-\varepsilon})$, 复杂路由需要至少 $cn^{\frac{\alpha-2\varepsilon}{2(\alpha+2)}}$ 步才能到达目标 t 。用类似的证明方法可以得到期望的结论。

有关复杂路由时间的下界对于分散式的路由选择算法依然成立, 这里本文提供了对于 $\alpha = 2$ 情况的证明。

令 \mathcal{A} 为所有的确定性分散式路由算法的集合, \mathcal{G} 为所有可能的基于 $\alpha = 2$ 的 *Kleinberg* 小世界网络的集合, π 是生成的 *Kleinberg* 小世界网络在 \mathcal{G} 上的分布。进一步定义 $T(A, G)$ 为分散式路由算法 $A \in \mathcal{A}$ 在小世界网络 $G \in \mathcal{G}$ 上的复杂路由时间。因为随机算法是在确定性算法上的一个分布, 我们只需要证明对于任意的在确定性算法集合 \mathcal{A} 上的分布 μ ,

$$\Pr_{G \sim \pi}(\mathbb{E}_{A \sim \mu}(T(A, G)) = \Omega(n^{1/4})) = 1 - O(\frac{1}{\log n}). \quad (5.2)$$



这里的概率是相对于 G 遵从的分布 π , 期望 $\mathbb{E}_{A \sim \mu}(T(A, G))$ 是对于随机算法 A 遵从的分布 μ 。

从前面的证明可以得出

$$\forall A \in \mathcal{A}, \Pr_{G \sim \pi}(T(A, G) = \Omega(n^{1/4})) = 1 - O\left(\frac{1}{\log n}\right).$$

因此存在不依赖于算法 A 的常数 $c_1, c_2 > 0$, 有

$$\forall A \in \mathcal{A}, \Pr_{G \sim \pi}(T(A, G) \geq c_1 n^{1/4}) \geq 1 - \frac{c_2}{\log n}.$$

由于这个不等式对于任意的 $A \in \mathcal{A}$ 都成立, 那么对于 \mathcal{A} 上的分布 μ

$$\Pr_{\substack{G \sim \pi \\ A \sim \mu}}(T(A, G) \geq c_1 n^{1/4}) \geq 1 - \frac{c_2}{\log n}. \quad (5.3)$$

再观察等式

$$\Pr_{G \sim \pi}(\mathbb{E}_{A \sim \mu}(T(A, G)) \geq \frac{c_1}{2} n^{1/4}) \geq 1 - \frac{3c_2}{\log n}, \quad (5.4)$$

这是等式 (5.2) 的等价版本。下面用反证法证明, 假设等式 (5.4) 不成立, 那么

$$\Pr_{G \sim \pi}(\mathbb{E}_{A \sim \mu}(T(A, G)) < \frac{c_1}{2} n^{1/4}) \geq \frac{3c_2}{\log n}.$$

令 $\mathcal{G}_1 \subseteq \mathcal{G}$ 为满足 $\mathbb{E}_{A \sim \mu}(T(A, G)) < \frac{c_1}{2} n^{1/4}$ 的所有小世界网络 G 的集合, 于是可以得到 $\Pr_{G \sim \pi}(G \in \mathcal{G}_1) \geq \frac{3c_2}{\log n}$ 。对于任意固定的小世界网络 $G \in \mathcal{G}_1$, 由马尔可夫不等式 (Markov's inequality $\Pr(|X| \geq a) \leq \frac{E(|X|)}{a}$) 可得

$$\forall G \in \mathcal{G}_1, \Pr_{A \sim \mu}(T(A, G) \geq c_1 n^{1/4}) \leq \frac{\mathbb{E}_{A \sim \mu}(T(A, G))}{c_1 n^{1/4}} < \frac{1}{2}.$$

因此对于任意的 \mathcal{G}_1 中的元素 G , $\Pr_{A \sim \mu}(T(A, G) < c_1 n^{1/4}) \geq 1/2$ 都成立, 而 \mathcal{G}_1 的元素出现的概率至少为 $\frac{1}{2}$ 。

$$\Pr_{\substack{A \sim \mu \\ G \sim \pi}}(T(A, G) < c_1 n^{1/4}) \geq \frac{1}{2} \cdot \frac{3c_2}{\log n} = \frac{1.5c_2}{\log n},$$

这与不等式 (5.3) 相矛盾, 假设不成立, 等式 (5.4) 和等式 (5.2) 成立。因此在 $\alpha = 2$ 时, 对于任意的随机路由算法 $A \sim \mu$, 期望的复杂路由时间均有 $\Omega(n^{1/4})$ 的下界。 \square



5.3 讨论和扩展

用上一节的方法,可以得到 k -复杂路由相同的下界。为了保证复杂路由能够找到目标,需要设置 Kleinberg 小世界网络的参数为 $p = q = k$, 同时种子节点的个数为 k 。因为结果和 2-复杂路由一样,这里就不在列出定理。

接下来本文把结果扩展到每一步允许感染多个 (m 个) 节点的复杂路由。当 $m = 1$ 时,这正是我们前一节在定理5.1讨论的复杂路由。当不限制 m 的大小时,复杂路由就演变为复杂传播。 m 可以被看作是连结复杂路由和复杂传播的变量,本文尝试去探索多大的 m 可以把复杂路由的时间拉低到复杂路由的水平。

定理 5.2: 对于任意的分散式路由策略 (算法), 在 $\alpha = 2$ 的有向 Kleinberg 小世界网络中, 每一步可以允许感染 m 个节点的 k -复杂路由的路由时间有很高的概率 $1 - O(\frac{1}{\log n})$ 路由时间为 $\Omega(\frac{n^{1/4}}{m})$, 期望路由时间为 $\Omega(\frac{n^{1/4}}{m})$ 。

证明 设 S 为当前已感染节点的集合。在下一步,通过 S 发出的有向边我们最多可以感染 m 个节点。但是考虑一步只允许感染一个节点的复杂路由,如果一步只去感染一个节点,但是分成 m 步去感染。在每一步感染完成后,路由算法都会了解到新感染节点的有向边,也就是有更多的机会去感染其他节点。不难看出每步感染一个节点分成 m 步去感染要比一步感染 m 个节点的路由更有效率。因此如果每步感染一个节点的复杂路由需要 T 个时间步骤来找到目标,每步允许感染 m 个节点的复杂路由至少也需要 $\frac{T}{m}$ 才能完成。所以每步允许感染 m 个节点的复杂路由的路由时间为 $\frac{T}{m} \cdot (1 - O(\frac{1}{\log n})) = \Omega(\frac{T}{m})$ 。证毕。 \square

从上面这个定理可以得知为了得到路由时间为 n 的多项式对数级的复杂路由,每一步允许感染的节点个数 m 需要很大,至少为 $m = n^{1/4} / \log^{O(1)} n$ 。

5.4 本章小结

本章提出了 Kleinberg 小世界网络中复杂路由的概念,并对这一信息扩散现象进行了研究。首先给出最简单的复杂路由的结论并给出了完整了理论证明,然后把结论扩展到 k -复杂路由,并且建立起复杂路由和复杂传播的关系,研究了每步允许感染 m 个节点的复杂路由。



6 总结与展望

6.1 工作总结

本章总结了论文取得的主要结论和成果，然后与已有结果进行对比。本文主要研究了 Kleinberg 小世界网络中复杂传染的现象，解决了去年遗留的开放性问题^[15]。文中大部分的结论是基于阈值为 2 的复杂传染病，在每一章的最后，把结论扩展到阈值为 k 的情况。表格6.1总结了本文的结果并且与之前有关复杂传染的结论对比，这里的结论都是基于有 n 个节点的二维 Kleinberg 小世界网络上阈值为 2 的复杂传染。

首先，本文解决了^[15]中提出的有关复杂传播的问题：对于 $\alpha \in (0, 2)$ 的具有 n 个节点的小世界网络，Ghasemiefteh 等人给出的传播时间的上界 $O(n^{\frac{2-\alpha}{5-2\alpha}} \log^{\frac{3}{2}} n)$ 和下界 $\Omega(\log n / \log \log n)$ 有指数量级的差距。本论文给出了 α 在区间 $(0, 2)$ 时传播时间多项式形式的下界（参见表格6.1中有关复杂传播结果的第四行），缩小了这个上下界的差距。

表 6.1 传播时间和路由时间的结果对比

α	$[0, \frac{1}{5})$	$[\frac{1}{5}, \frac{4}{5}]$	$(\frac{4}{5}, 2)$	2	$(2, 3]$	$(3, 4)$	4	$(4, \infty)$
简单传播（直径）	$\Theta(\log n)^*$			$O(\log^\beta n)^+$?	$\Omega(n^{\frac{\alpha-4}{2(\alpha-2)}})^+$
简单路由	$\Omega(n^{\frac{2-\alpha}{6}})^\S$			$O(\log^2 n)^\S$	$\Omega(n^{\frac{\alpha-2}{2(\alpha-1)}})^\S$			
复杂传播	$\Omega(n^{\frac{1-\alpha}{4}})$	$\Omega(n^{\frac{1}{5}})$	$\Omega(n^{\frac{2-\alpha}{6}})$	$O(\log^{3.5} n)^\#$?	$\Omega(n^{\frac{\alpha-3}{2\alpha}})$		
复杂路由	$\Omega(n^{\frac{1}{\alpha+2}})$			$\Omega(n^{\frac{1}{4}})$	$\Omega(n^{\frac{\alpha}{2(\alpha+2)}})$			

注 1：复杂传播和复杂路由的结果是阈值为 2 的情况， α 是小世界模型的参数，表中黑体部分是本文的结果。

注 2：“*”表示 Martel 等人的结果^[13]，“+”表示 Nguyen 等人的结果^[14]，其中 $\beta = \log_{4/(\alpha+\varepsilon)} 2 + \frac{1}{\varepsilon}$ ，“§”表示 Kleinberg 的结果^[8]，“#”表示 Ghasemiefteh 等人的结果^[15]。

然后本文指出 $\alpha > 3$ 时，复杂传播时间期望的下界为 $\Omega(n^{\frac{\alpha-3}{2\alpha}})$ ，（以很高的概率 $1 - O(n^{-\varepsilon})$ ，传播时间至少为 $\Omega(n^{\frac{\alpha-3-\varepsilon}{2\alpha}})$ ）。结合 $\alpha \in (0, 2)$ 的结果，可以看出对于 $\alpha < 2$ 和 $\alpha > 3$ ，传播时间均有 n 的多项式形式的下界（ $\alpha \in (2, 3]$ 时的上下界仍然未知）。定性地看，复杂传播表现出和 Kleinberg 提出的分散式路由^[8]相似的性质（Kleinberg 的分散式路由当且仅当在最佳点 $\alpha = 2$ 时有 n 的多项式对数的上界）。这里我们对比简单传染病和复杂传染病的结果。可以观察到简单传染病的传播比较像从一个节点开始的广度优先搜索，因此简单传播时间对应着网络的直径。近期的研究^[13, 14]指出当 $\alpha < 4$ 时，Kleinberg 小世界网络的直径都是 n 的多项式对数级（见表格6.1的第二行）。从这可以



看出, 从简单传染病到复杂传染病 (k 从 1 跳变到 2), $\alpha \in [0, 2) \cup (3, 4)$ 时的传播时间完成了从 n 的多项式对数级到多项式级的巨大跳变, 出现了相变现象。这暗示了对于通常的 α (不包括最佳点 $\alpha = 2$), 复杂传染病的传播要比简单传染病的传播慢得多。

接下来本文在复杂传播的基础之上, 进一步研究了一种新的扩散现象: 复杂传播的路由, 这比较像^[8]中的分散式路由。在复杂路由中, 小世界网络的弱连接被认为是向边。本文定量地研究了从两个相邻的种子节点开始去感染一个网格上最远目标 t 需要的步数, 也就是路由时间。在每一步, 只有一个节点会被感染, 选择被感染节点的策略是分散式的, 也就是路由选择策略只能基于当前被感染的节点发出的强连接和弱连接。复杂路由比较像社交网站上最近提出的一个应用: 主动交友, 主动交友是指在像 Facebook、人人网等社交网站上通过添加一些中间好友来增加目标接受自己好友请求的概率^[22]。

本文指出与复杂传播不同, 对于所有的 α , 复杂路由时间均有 n 的多项式的下界 (参见表格 6.1 的第五行)。与复杂传播的结果比较可以发现, 在最佳点 $\alpha = 2$ 处复杂传播的传播时间上界为 $O(\log^{3.5} n)$, 而复杂路由的期望路由时间具有 $\Omega(n^{\frac{1}{4}})$ 的下界。这个指数级的差别揭示了复杂传播和复杂路由的内在机制差别: 复杂传播过程中每一步会尽可能多地去感染节点, 而复杂路由则只能一步感染一个节点。本文进一步指出如果改进复杂路由, 即使允许在路由的每一步感染 m 个节点, 路由时间仍然具有 $\Omega(\frac{n^{\frac{1}{4}}}{m})$ 的下界。这也就是说如果想要得到多项式对数级的复杂路由时间, m 至少需要为 $\Omega(\frac{n^{\frac{1}{4}}}{\log^c n})$ 。因为 Kleinberg 的分散式路由^[8]对应着简单传染病的路由, 因此还可观察到简单路由和复杂路由间也有量级的差别。简单路由在最佳点 $\alpha = 2$ 具有上界为 $O(\log^2 n)$ 的路由时间, 而对于所有的 α , 复杂路由时间均有 n 的多项式的下界。

本文的结论为复杂传染病比简单传染病更难传播这一观点^[12] 提供了强有力的理论支持, 也揭示了复杂路由甚至比复杂传播更难传播。

6.2 工作展望

本文主要对小世界网络中的复杂传播和复杂路由进行了研究, 但是结论仍有一些缺陷, 仍然有很多方面需要改进和进一步研究。

首先, 在后续的工作中希望进一步完善已有的结论, 尤其是给出 2-复杂传播时间在 $\alpha \in (2, 3]$ 时的下界。蒙特卡洛模拟结果显示 $\alpha \in (3, +\infty)$ 复杂传播的时间远大于 $\alpha \in [0, 2)$ 的复杂传播时间, 而本文给出的下界中 $\alpha \in (3, +\infty)$ 传播时间的下界却比 $\alpha \in [0, 2)$ 的下界弱, 这与模拟结果不符, 也说明 $\alpha \in (3, +\infty)$ 时传播时间的下界有待提高。



其次, 在研究复杂路由时, 为了更好地应用延迟选择原则, 本文采用了有向 Kleinberg 小世界网络模型。因为每个节点的弱连接平均数量仍为常数, 有向小世界网络和无向小世界网络中复杂路由的效率应该不会相差太多。后续的工作希望可以研究复杂路由在无向 Kleinberg 小世界网络中的性能。

最后, 复杂传播和复杂路由在 $\alpha = 2$ 处效率相差很大, 本文用一步允许感染 m 个节点的复杂路由建立起与复杂传播的关系, 但是复杂路由的效率仍然很低。未来可以研究在什么样结构的网络中复杂路由可以很快找到目标, 探讨复杂路由在其他小世界模型中的执行效率。为复杂路由寻找更广泛的应用场景也是未来需要进行的工作之一。



致谢

在完成本科学位论文之际，我在此向所有支持我、关心我的老师、师兄、同学和家人致以最衷心的感谢。是你们的帮助，让我能够顺利完成毕业设计。感谢北航软件工程研究所对我的支持，感谢 G305 实验室给我在这里的实习机会，让我能够在一个科研环境优越的地方研究学习。

首先十分感谢我的指导教师任爱华老师。自从我大三进入实验室以来，任老师在科研、学习、生活方面给了我非常多的帮助。任老师严谨科学的学术态度，使我受益良多，她在和我们日常交流中风趣幽默，也在潜移默化地教导我们如何做人如何做事。感谢任老师对我论文的选题、设计、完成以及论文修改等各个方面给予的悉心指导。我也要感谢答辩组的各位老师，刘超老师、吴际老师、王雷老师和蒋竞老师，你们在开题答辩和中期检查时都给提出了我非常重要的指导意见，让我对自己的工作有了更深刻的认识，得到了新的启示。

我还要感谢在微软亚洲研究院实习时对我毕业设计进行过指导的陈卫老师和中科院计算所的孙晓明老师和张家琳老师，每周和你们的讨论我都收益颇多，你们对我在论文写作部分给予了细致的指导。也是在与你们的讨论中我逐渐发现了自己的兴趣，在这里衷心地感谢你们对我的教导。

感谢在微软亚洲研究院一起实习的小伙伴，肖文聪、王安然、谭传奇、吴俣同学，在工作和学习时遇到问题一起讨论，你们陪伴我度过了大学的这段实习生活。同时也感谢微软的同事孙雷和郑书新同学，能和你们一起工作是我的荣幸。

感谢我的好朋友王裔鸣同学，毕设过程中我们互相鼓励，谢谢你的一路支持。也感谢我的舍友银一鑫、刘乃维和牛晨阳以及实验室的运明纯同学，是你们陪伴我走过了毕设的日子，谢谢你们对我的鼓励和支持。我还要感谢实验室的学长学姐们，王洪江师兄、丁子然师兄、徐博儒师兄、何芬学姐和陈云学姐，感谢你们对我毕业设计提出了宝贵的意见，耐心的解答我们的疑惑。

感谢陪我走过大学四年的辅导员和 1006 大班的同学们，和你们在一起的四年让我收获了很多人生财富。

最后要感谢我的家人对我的支持和鼓励，特别是父亲和母亲，你们的鼓励永远让我充满动力！



参考文献

- [1] Milgram S. The small world problem[J]. *Psychology Today*, 1967, 2(1):60–67.
- [2] Watts D. J., Strogatz S. H. Collective dynamics of ‘small-world’ networks[J]. *Nature*, 1998, 393(6684):440–442.
- [3] Albert R., Jeong H., Barabási A.-L. Internet: Diameter of the world-wide web[J]. *Nature*, 1999, 401(6749):130–131.
- [4] Granovetter M. S. The strength of weak ties[J]. *American Journal of Sociology*, 1973:1360–1380.
- [5] Granovetter M. Getting a job: A study of contacts and careers[J]. 1974.
- [6] Sola Pool I., Kochen M. Contacts and influence[J]. *Social Networks*, 1978, 1(1):5–51.
- [7] Bollobás B. Random graphs[M].[S.l.]: Springer, 1998.
- [8] Kleinberg J. The small-world phenomenon: an algorithm perspective[A]. *Proceedings of the thirty-second annual ACM Symposium on Theory of Computing*[C]. [S.l.]: [s.n.], 2000:163–170.
- [9] Newman M. E., Watts D. J. Scaling and percolation in the small-world network model[J]. *Physical Review E*, 1999, 60(6):7332.
- [10] Granovetter M. Threshold models of collective behavior[J]. *American Journal of Sociology*, 1978:1420–1443.
- [11] Kempe D., Kleinberg J., Tardos É. Maximizing the spread of influence through a social network[A]. *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*[C]. [S.l.]: [s.n.], 2003:137–146.
- [12] Centola D., Macy M. Complex contagions and the weakness of long ties[J]. *American Journal of Sociology*, 2007, 113(3):702–734.
- [13] Martel C., Nguyen V. Analyzing Kleinberg’s (and other) small-world models[A]. *Proceedings of the twenty-third annual ACM Symposium on Principles of Distributed Computing*[C]. [S.l.]: [s.n.], 2004:179–188.
- [14] Nguyen V., Martel C. Analyzing and characterizing small-world graphs[A]. *Proceedings of the sixteenth annual ACM-SIAM Symposium on Discrete Algorithms*[C]. [S.l.]: [s.n.], 2005:311–320.
- [15] Ghasemiesfeh G., Ebrahimi R., Gao J. Complex contagion and the weakness of long ties



- in social networks: revisited.[A]. ACM Conference on Electronic Commerce[C]. .[S.l.]: [s.n.] , 2013:507–524.
- [16] Fraigniaud P., Giakkoupis G. The effect of power-law degrees on the navigability of small worlds: [extended abstract][A]. Proceedings of the 28th Annual ACM Symposium on Principles of Distributed Computing[C]. .[S.l.]: [s.n.] , 2009:240–249.
- [17] Fraigniaud P., Giakkoupis G. On the searchability of small-world networks with arbitrary underlying structure[A]. Proceedings of the 42nd ACM Symposium on Theory of Computing[C]. .[S.l.]: [s.n.] , 2010:389–398.
- [18] Kempe D., Kleinberg J., Tardos É. Influential nodes in a diffusion model for social networks[A]. Proceedings of the 32nd international conference on Automata, Languages and Programming[C]. .[S.l.]: [s.n.] , 2005:1127–1138.
- [19] Chen W., Wang Y., Yang S. Efficient influence maximization in social networks[A]. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C]. .[S.l.]: [s.n.] , 2009:199–208.
- [20] Chen W., Wang C., Wang Y. Scalable influence maximization for prevalent viral marketing in large-scale social networks[A]. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining[C]. .[S.l.]: [s.n.] , 2010:1029–1038.
- [21] Chen N. On the approximability of influence in social networks[A]. Proceedings of the nineteenth annual ACM-SIAM Symposium on Discrete Algorithms[C]. .[S.l.]: [s.n.] , 2008:1029–1037.
- [22] Yang D.-N., Hung H.-J., Lee W.-C., et al. Maximizing acceptance probability for active friending in online social networks[A]. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C]. .[S.l.]: [s.n.] , 2013:713–721.
- [23] Motwani R., Raghavan P. Randomized Algorithms[M].[S.l.]: Cambridge Univ Press, 1995.