arXiv:2304.04480v1 [cs.CR] 10 Apr 2023

# 局部交互主体网络中高度组织化社群的存在性

V.利亚古[1,4], P.E.纳斯图[5,6],P. 斯皮拉基斯[2],Y.C. 斯塔马蒂乌[1,3]

[1] 帕特雷大学计算机技术研究所与出版社——“丢番图”，希腊帕特雷大学校园，邮编 26504[2] 英国利物浦大学计算机科学系，以及帕特雷大学计算机工程与信息学系，希腊帕特雷，邮编 26504[3] 帕特雷大学工商管理系，希腊帕特雷，邮编 26504[4] 约阿尼纳大学信息学与电信系，希腊阿尔塔州科斯塔基奥伊，邮编 47100[5] 爱琴海大学数学系，应用数学与数学建模实验室，希腊萨摩斯岛[6] 佛罗里达大学应用优化中心，美国佛罗里达州盖恩斯维尔 电子邮箱：liagkou@cti.gr, pnastou@aegean.gr, P.Spirakis@liverpool.ac.uk, stamatiu@ceid.upatras.gr

2023年4月11日

**摘要**。本文研究相关主体网络中高度组织化结构自发涌现或有目的形成的现象。我们表明，大型有序结构的形成需要规模随结构大小呈指数增长的网络。我们的方法基于将网络视为有限长度字符串时的柯尔莫哥洛夫复杂度，或描述复杂度。我们运用这一方法研究基于谢尔宾斯基图的简单有序层级结构的涌现或形成，并证明了一类拉姆齐定理，该定理限定了包含谢尔宾斯基图作为子图的柯尔莫哥洛夫随机图中的顶点数量。此外，我们表明，谢尔宾斯基图包含其顶点之间的紧密连接的关系；当其顶点中的主体参与以双人博弈为模型的成对互动时，这些关系有助于信息的快速传播与学习。最后，我们将研究结果推广到任何具有简洁表示的有序结构。特别是，我们的工作可用于研究与网络安全相关的问题，通过识别能够促成或阻止形成足够大的内部子网络的条件，这些内部子网络具有恶意的共同目标，旨在接管网络或扰乱其运行。

**关键词：** 有序威胁结构、网络威胁、谢尔宾斯基三角形、谢尔宾斯基图、Kolmogorov 复杂度、拉姆齐理论。

## 1 Introduction

本文探讨某些配置或模式在主体演化网络（例如社交网络）中形成或涌现的问题。在本文语境中，我们将这些配置或模式称为高度组织化结构，并结合图中的拉姆齐数与*Kolmogorov* 复杂度的概念进行研究。我们的工作不依赖随机图或渐近分析技术，而是聚焦于有限图，可用于研究有序结构何时以及如何在有限规模的交互主体网络中出现，并回答如下问题：(i) 给定一个特定规模的交互主体网络，其中能够形成多大的有序子网络，使共同目标得以扩散并最终席卷整个网络？(ii) 特定组织结构的子群能否出现在给定规模的网络中？(iii) 交互主体网络需要多大，才会容易受到大规模、紧密交互的恶意主体子网络形成的威胁？

广义而言，拉姆齐理论指的是这样一类数学命题：某种特定结构（例如字符串、图、数列等）必然包含一个规模较大且高度有序的子结构。这类命题的例子来自多个学科，包括数理逻辑、数论、分析和图论。数论中最著名的此类命题之一由 van der Waerden 于 1927 年证明（见 [15]），内容如下：对于任意给定的正整数 $r$ 和 $k$，都存在某个正整数 $n$，使得如果整数 $\{1, 2, \ldots, n\}$ 用 $r$ 种不同的颜色着色，那么至少存在 $k$ 个构成等差数列的同色整数。满足条件的最小 $n$ 称为 Van der Waerden 数 $W(r, k)$。本文关注的是图的拉姆齐理论。该研究领域由 Ramsey 于 1930 年在 [12] 中开创，他在其中陈述并证明了所谓的 Ramsey 定理：对于任意图 $H$，都存在一个自然数 $n$，使得对 $K_n$（即具有 $n$ 个顶点的团）的边进行双色染色时，$K_n$ 必定包含 $H$ 的一个单色副本作为子图，但该子图不一定是诱导子图。满足条件的最小值 $n$ 称为 $H$ 的 Ramsey 数，记作 $r(H)$（如果 $H = K_t$，则简记为 $r(t)$）。本文研究谢尔宾斯基图这一类图的 Ramsey 数。这类图的顶点最大度有界。这一性质使得这些图的 Ramsey 数存在关于 $H$顶点数的线性上界。这是 Chv´atal、R¨odl、Szemer´edi 和 Trotter 在 [4] 中的一个更一般结果的推论：该结果指出，若 $H$ 是一个具有 $n$ 个顶点且最大度为 $\Delta$的图，那么其 Ramsey 数

$r(H)$以 $c(\Delta)n$ 为上界，其中 $c(\Delta)$是仅取决于 $\Delta$的某个常数。也就是说，有界度图的 Ramsey 数随顶点数线性增长。此外，诱导Ramsey 数也存在一个线性上界。诱导 Ramsey 数的定义与 Ramsey 数相同，但现在 $H$ 的单色副本必须构成一个诱导子图。这是对子图更严格的要求：子图 $H$ 中存在边，当且仅当这些边也是 $H$中的边。较大图中 $H$ 的顶点之间可能存在的其他边则不能存在；而在一般的子图概念中，这种情况是允许的。诱导*Ramsey*定理指出，对于每个图 $H$ ， $r_{\mathrm{ind}}(H)$都存在（例如参见 [6]第 9.3 章）。

另一方面，*Kolmogorov* 复杂度对有限对象（通常是对图等有限结构进行建模的比特序列）的定义是：在不丢失信息的前提下，能够压缩该对象所需的最少比特数。也就是说，压缩后的字符串可以通过某种算法在参考机上恢复，而参考机通常是通用图灵机（见 [9]）。请注意，在 Kolmogorov 复杂度的语境中，“ 复杂 ”一词指的更接近于缺乏组织、模式和“精细性”，而非日常语言中用来形容具有高度组织性和精细结构细节的对象或实体的含义。因此，Kolmogorov 复杂度高的对象，实际上按通常意义来说复杂度较低，因为它们具有随机性的所有性质；而 Kolmogorov 复杂度低的对象则缺乏随机性，具有内部结构组织和许多规律性。因此，用柯尔莫哥洛夫复杂度理论的术语来说，它们可以被简洁地描述或压缩。关于这一复杂度概念，它由安德烈·柯尔莫哥洛夫、雷·所罗门诺夫和格雷戈里·蔡廷分别独立提出并发展（[8,10,3]）。这一概念讨论的是有限对象的复杂度，与 NP完全性等经典复杂性理论不同；后者讨论的是有限对象的无限集合的复杂度，即复杂性理论术语中的语言。

基于拉姆齐理论和 Kolmogorov 复杂度的论证，本文研究演化主体网络中高度组织化结构（如组织、人际网络和社会模式）的存在或形成现象。作为一个具体案例，我们聚焦于谢尔宾斯基图这一类图，它们还具有紧密连接的重要性质；也就是说，基于这些图相互连接的主体具有很强的“凝聚力”。紧密图族在图中的扩散过程（见 [13]）中发挥着重要作用，例如在.....

作为在大量主体中共同且达成一致的创新或思想的涌现.....

## 2 图的紧密性性质

对智能体交互及其长期演化和基于此类模型的性质进行研究，可以揭示有关社会结构动态、其扩展方式，以及这些动态必然促成何种组织化子结构涌现的重要信息。例如，本文将证明，在特定条件下，可以在包含交互智能体的大型结构中找到具有特定“成员紧密性”性质的大型子结构。本文还将证明，随着网络扩展，大型紧密连接的规律性子结构会涌现，形成局部的规律性与社会组织“孤岛”，即使主体网络在不可预测的（即随机的）全局方式下演化也是如此。正如 Young 在 [13], 中所指出的，紧密连接的主体结构在传播观念与信念方面发挥着重要作用，而这些观念与信念起初仅由一小群交互智能体持有。

Young 的数学结果考察了某种观念或观点如何在交互智能体群体中传播；该群体的社会结构具有某种“连贯性”性质，其成员则通过双人“创新接受”博弈进行互动。

在 [13], Young 中定义了主体网络的一个参数，这些网络被建模为图，该参数描述了它们的“连贯性”，以及它们对于局外人的观点抵御待传播创新的脆弱性。一个主体群体$S$ （即给定图 $G$ 中没有孤立顶点的节点）如果对于每个 $S' \subseteq S$， $S' \neq \emptyset$，在为 $r$ 取适当值的情况下（见下文讨论），以下条件成立，则称为紧密连接的：

$$\min_{S' \subseteq S} \frac{d(S', S)}{\sum_{i \in S'} d_i} \geq r \tag{1}$$

其中， $d(S', S)$是 $S'$ 的内部连接，即 $\{i, j\}$ 条边的数量，其中 $i \in S'$ 且 $j \in S$ ，而 $d_i$ 是主体 $i$ 所拥有的边总数（即其在图中的度）。由不等式（1）显然可知，对于紧密连接的群体 $S$ 的每个子集，内部度与总度之比至少为 $r$。直观地说，要使智能体群体中的这一比率较大，相对而言，我们需要许多内部连接和很少的外部连接。

给定一个正整数 $k$ 和一个实数 $0 \leq r \leq 1/2$，若每个个体都属于一个基数为

至多为 $k$ 的群体 $k$，且该群体按照不等式（1）的定义是 $r$-紧密连接的，则称图（，$0 \leq r \leq 1/2$ ）为紧密连接的。最后，如果对每个 $0 \leq r \leq 1/2$ ，都存在一个可能依赖于 $r$ 的整数 $k$，使得该类中的每个图都是（$r, k$）-紧密连接的，则称图的类是紧密连接的 。

此外，该群体中的主体会定期参与双人博弈，在这种博弈中，成对的交互智能体相互竞争以获得收益（非零和博弈）。根据该博弈，每当两个主体进行交互时，他们会根据各自选择所能获得的收益，独立地选择是否采用该创新。假设该博弈具有一个风险占优纳什均衡，在这一均衡中两名玩家都采用该创新（详情见 [13] ）。

基于这两个要素，Young 证明，对于一类紧密相连的智能体图，所有社区成员最终都会采用这项创新，即博弈中的风险占优纳什均衡，而且所需的交互次数（“时间”）有界，且不取决于社区规模。换言之，如果社区成员的所有子集之间都存在强有力的成对边，同时又与圈外人联系较弱（圈外人甚至可能反对采用这项创新），那么一群发起者最终将能够说服所有群体成员采用这项创新。

## 3 基于谢尔宾斯基三角形的群体编队

在本节中，我们描述一类基于谢尔宾斯基三角形分形的图，并证明它构成一个紧密连接的图族。

**定义 1（层级为 $l$ 的谢尔宾斯基三角形垫片）**。给定一个整数 $l, l \geq 1$，我们将层级为 $l$ 的谢尔宾斯基三角形垫片，或简称层级为 $l$ 的谢尔宾斯基三角形，定义如下：对于 $l = 1$ ，谢尔宾斯基三角形是一个等边三角形；而对于 $l > 1$，层级为 $l$ 的谢尔宾斯基三角形由三个层级为 $l-1$ 的谢尔宾斯基三角形副本在各自的角上连接而成。

在图 1 中，我们可以看到前几个谢尔宾斯基三角形，分别对应 $l = 1$，2，3，4，5，6 和 7。基于谢尔宾斯基三角形，我们可以定义一个对应的图族，称为谢尔宾斯基图。

**定义 2（谢尔宾斯基图）**。层级为 $l$、$l \geq 1$ 的谢尔宾斯基图，记为 $S_l$ ，构造如下：若 $l = 1$ ，则将三个顶点及边替换后，即构成层级为 *1* 的谢尔宾斯基图。



**图 1**。 谢尔宾斯基三角形，对于 $l = 1$，2，3， 4，5， 6 和 7，分别。

层级*1*的谢尔宾斯基三角形及其图顶点和边；对于$l > 1$，层级*1*的谢尔宾斯基图由三个层级 $l-1$ 的谢尔宾斯基图副本构成，通过识别它们中分别对应于层级 $l-1$的谢尔宾斯基三角形各个角的顶点。

在图 2 中，我们可以看到谢尔宾斯基图 $S_1, S_2$ 和 $S_3$。特别是，

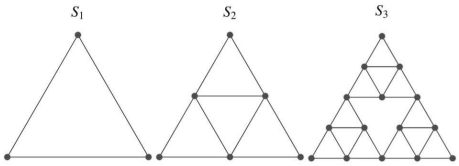


**图 2.** 谢尔宾斯基图 $S_1, S_2$ 和 $S_3$。

在图3中，我们看到层级4的谢尔宾斯基图 $S_4$由三个层级3的谢尔宾斯基图 $S_3$副本组成，后者显示在虚线框内。不难看出，以下性质成立

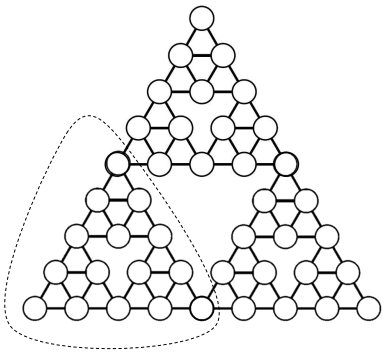

**图 3**。 层级 4 的谢尔宾斯基图。

对于 $S_l$:

– 对于 $l > 1$ 的每个值， $S_l$ 的所有顶点的度均为 4，除了三个度为 2 的顶点（即“角点”）。– $S_l$ 的顶点数为 $n_l = \frac{3}{2}(3^{l-1} + 1)$。– $S_l$ 的边数为 $m_l = 3^l$。

度和连通性有助于满足不等式（1）。直观地说，拥有过多度数较大的节点，会使一群主体暴露于过多的外部干扰之下。这会对紧密联系性质产生负面影响。我们证明如下：

**定理1**。 谢尔宾斯基图的族具有紧密连接的性质。

证明。 令 $S_l = (V, E)$ 为一个层级为 $l, l \geq 1$ 的谢尔宾斯基图。令 $S$ 为其顶点集合 $V$ 的一个子集，使得 $|S| = k$，且 $S'$ 是 $S$ 的一个非空子集。 $S$ 和 $S'$ 中的每个顶点的度均为 2 或 4。此外，令 $i, 0 \leq i \leq 3$ 表示 $S'$ 中度为 2 的顶点数。我们计算不等式（1）中分数的分子和分母。

若以 $\alpha, 0 \leq \alpha \leq 1$ 表示 $S'$ 中的顶点与 $S - S'$ 中的顶点之间的边的比例，则对于不等式（1）中分数的分子，我们有

$$d(S', S) = \alpha[4(k-i) + 2i] + (1-\alpha)\ \frac{4(k-i)+2i}{2} \tag{2}$$

$$= \alpha[4(k-i) + 2i] + (1-\alpha)\ [2(k-i) + i]. \tag{3}$$

由于在计算（3）中的 $d(S', S)$ 时， $S'$ 中顶点之间的边只能计数一次。至于不等式（1）中的分母，我们有

$$\sum_{i \in S'} d_i = 4(k-i) + 2i. \tag{4}$$

由（3）和（4）可得

$$\frac{d(S', S)}{\sum_{i \in S'} d_i} = \frac{\alpha[4(k-i) + 2i] + (1-\alpha)\ [2(k-i) + i]}{4(k-i) + 2i}. \tag{5}$$

经过一些直接的运算，(5) 化简为 $\frac{1+\alpha}{2}$ 。

因此，不等式（1）对任意 $r$ 和 $k$均成立。我们得出，对于 $l \geq 1$， $S_l$，即谢尔宾斯基图，构成一个紧密连接的图族。 □

## 4 交互智能体网络中高度组织化结构的存在

在本节中，我们探讨交互主体网络中出现高度组织化结构所需的条件。这些结构

可以代表任何具有共同目标或共同信念的个体组织化共同体。宽泛地说，这些条件取决于结构是否具有某些可识别的规律性特征，以便对其进行简洁地描述。这些规律性和简洁性条件刻画了“有序结构”这一概念。具体而言，我们将组织化且具有规律性的结构限定为谢尔宾斯基图这一类，并且它们还具有紧密连接的性质（见第 3 节）。

更具体地说，我们将研究在足够大的图中，作为子图的谢尔宾斯基图是否存在这一问题；这些图可以对不断演化的交互智能体、社会网络和社会进行建模。本节将运用 Kolmogorov 复杂度和拉姆齐理论中的技术。接下来，我们将简要介绍*Kolmogorov* 复杂度 和 拉姆齐理论的主要定义及一些有用结果。随后，我们将结合这两种理论，探究在以图为模型的演化结构中，促成或阻碍有组织的子结构涌现的条件。

### 4.1 Kolmogorov 复杂度

通俗地说（参见 [8,10,3]），（二进制）字符串 $x$ 的 *Kolmogorov* 复杂度是其 最短算法描述的 长度。换句话说，*Kolmogorov* 复杂度记作 $C(x)$，对于有限字符串 $x$ 而言，是指将其编码为比特二进制序列的、能够在不接受任何输入的情况下生成 $x$ 作为输出的 最短程序（或一般意义上的图灵机）的长度。同样地，给定 $y$时 $x$ 的 条件Kolmogorov 复杂度，记作 $C(x|y)$，是指在以 $y$ 作为输入的情况下生成$x$ 作为输出的最短程序的长度。可以证明， $C(x)$在某种意义上是通用的，因为除一个固定的加性常数之外，它不依赖于所选择的编程语言或图灵机模型；该常数取决于这一选择，但不取决于 $x$。

在本文中，我们关注的是图。我们可以通过以下方式用字符串对图进行编码，从而运用 Kolmogorov 复杂度的概念及其性质（例如，参见 [9]）：

**定义 3**。每个标记图 $G=(V,E)$ 若有 $n$ 个节点，都可以用一个长度为 $\binom{n}{2}$ 的二进制字符串 $E(G)$ 表示（在图同构意义下）。我们只需假定对一个有 $n$ 个节点的图中的 $\binom{n}{2}$ 条可能边进行固定排序，例如按字典序排列，并令字符串中的第 $i$ 位表示第 $i$ 条边是否存在（*1* 表示存在，*0* 表示不存在）。反过来，每个长度为 $\binom{n}{2}$ 的二进制字符串都编码一个有 $n$ 个节点的图。因此，我们可以将每个这样的图与其二进制字符串表示对应起来。

**定义 4**。一个有 $n$ 个节点的标记图，其随机性缺陷至多为 $\delta(n)$，若满足以下条件，则称其为 $\delta(n)$-随机：

$$C(E(G)|n) \geq \binom{n}{2} - \delta(n). \tag{6}$$

此外，以下结论成立（例如，见 [9]）：

**引理 1.** 至少有 $1-1/2^{\delta(n)}$ 比例的所有定义在 $n$ 个节点上的标记图 $G$ 是 $\delta(n)$-随机的。特别地，对于 $\delta(n) = \log n$，所有定义在 $n$ 个顶点上的图中，有$(\ -1-\frac{1}{n})$的分数是 $\log n$-随机的。

**定义 5.** 设 $G=(V,E)$ 是一个定义在 $n$ 个节点上的标记图。考虑一个定义在 $k$ 个节点 $\{1,2,\ldots,k\}$ 上的标记图 $H$。$k$ 个节点的每个子集在$G$中诱导出一个子图 $G_k$，该子图属于 $G$。当通过重新标记节点 $i_1 < i_2 <$得到 $H$ 时，子图 $G_k$ 是 $H$ 的一个有序标记出现
$\cdots < i_k$的 $G_k$为1、*2</style*、、.....、 $k$。

### 4.2 拉姆齐理论

我们首先给出拉姆齐理论中集合的 $r$ 染色的定义。

**定义 6.**（$r$-着色）设 $S$ 为一个集合，且 $r \in Z_+$。$S$ 的 $r$-着色是一个函数 $f$： $S \to \{1,2,\ldots,r\}$。

正如引言中所讨论的，拉姆齐理论关注的是在足够大的图中出现某些模式的问题。例如，拉姆齐理论最初源于以下问题：给定一个图 $H$，确定 Ramsey数 $r(H)$，其定义为最小的 自然数 $n$，使得任意 对 $E(K_n)$ 进行的二着色都包含 $H$ 的单色副本。在 [5]中证明了以下结论：

**定理2**。存在常数 $c$，使得任意图 $H$ 具有 $k$ 个顶点且最大度为 $\Delta$ 时都满足

$$r(H) \leq k\ 2^{c\Delta \log \Delta}.$$

在本文中，我们关注给定图 $H$ 的诱导拉姆齐数，其定义如下：

**定义 7.**（诱导 *Ramsey* 数）如果图 $H$ 是图 $H$ 的诱导子图，则 $V(H) \subset V(G)$ 成立；并且，$H$ 中的两个顶点相邻，当且仅当它们在 $G$中相邻。诱导 *Ramsey* 数 $r_{ind}(H)$定义为满足以下条件的最小整数：存在一个有 $r_{ind}(H)$个顶点的图 $G$，使得对 $G$ 的边进行任意二着色后，都包含 $H$ 的一个诱导单色副本。

请注意， $H$ 的诱导单色副本同时也是 $H$ 的一个诱导副本，即按通常图论意义而言的诱导子图，而无论 $G$的边是什么颜色。

已有许多结果给出了稀疏图的诱导拉姆齐数上界。例如，Beck 在 [1] 中研究了 $H$ 为树的情形。此外，Haxell、Kohayakawa 和 Łuczak [7] 证明，长度为 $k$ 的环的诱导拉姆齐数关于 $k$呈线性增长。更进一步， Łuczak 和 R¨odl [11]证明，最大度有界的图的诱导 Ramsey 数至多关于其顶点数呈多项式增长，从而解决了 Trotter 提出的一个猜想。更确切地说，他们证明了以下结论：

**定理3.** 对于每个整数 $d$，存在一个常数 $c_d$，使得每个有 $k$ 个顶点且最大度至多为 $d$ 的图 $H$ 都满足 $r_{ind}(H) \leq k^{c_d}$。

该证明给出的 $c_d$ 上界是一个高度与 $d^2$成正比的 2 的幂塔。由于谢尔宾斯基图的最大度为 4，因此将定理3应用于 $d = 4$，可立即得到以下推论：

**推论1**。 设 $H = S_l$ 是一个层级为 $l$ 、含有 $n_l$个顶点的带标号谢尔宾斯基图。那么 $r_{ind}(S_l) \leq n_l^{c_d}$，其中 $c_d$ 是一个正的常数，且与 $n_l$ 无关，因此也与 $l$无关。

### 4.3 演化图中有组织子图的涌现

尽管随机图（例如参见 [2]）是证明极限性质（即当图的大小渐近增长时成立的性质）的有力工具，但其主要局限在于，对于固定大小 $n$ 的特定有限图实例，它无法提供任何信息。另一方面，柯尔莫哥洛夫复杂度理论着重研究特定的有限对象（在此即图）。基于该理论，我们可以证明以下结论：

**定理4**。 设 $S_l$ 是一个含有 $n_l$ 个顶点的带标号谢尔宾斯基图。设 $G$是一个含有 $n$ 个顶点的带标号不可压缩图，并且包含 $S_l$ 作为诱导子图。那么 $n \geq 2^{\frac{n_l-1}{2}}$。

证明。 设 $G$ 为一个带标签的不可压缩图，其具有 $n$ 个顶点，且其编码 $E(G)$ 作为二进制字符串（见定义 3）的长度为 $l(E(G))$。由于 $G$ 是不可压缩的，因此有

$$C(E(G)|n, P) \geq \frac{n(n-1)}{2} \tag{7}$$

其中 $P$ 是一个程序，能够根据值 $n$ 以及下文所述的另一种编码 $E'(G)$，重构 $E(G)$。

由于根据我们的假设， $G$ 包含 $S_l$ 作为一个诱导子图，我们可以通过以下方式从 $E(G)$构造另一种编码 $E'(G)$来描述 $G$：

1.我们在编码 $E(G)$中加入对 $G$ 的谢尔宾斯基图子图的描述。为了描述图 $G$ 的 $n_l$ 个顶点中的这样一个图，我们需要 $\log \binom{n}{n_l}$ 位来表示这些 $n_l$ 个顶点的子集合，并需要 $\log n_l!$ 位来表示提供谢尔宾斯基图结构的特定排序$i_1 i_2 \dots i_{n_l-1} i_{n_l}$，正如我们在图2的示例图中所看到的那样；也就是说，我们在 $n_l$ 个顶点上构造谢尔宾斯基图，并从上到下、从左到右为它们编号（例如，关于谢尔宾斯基图的“顶部”和“底部”的指示，请参见图 2）。总之，要描述这个谢尔宾斯基图子图，我们至多需要 $\log \binom{n}{n_l} + \log n_l!$ 位，并且由于$\binom{n}{n_l} \leq \frac{n_l^n}{n_l!}$，我们至多需要

$$\log \frac{n_l^n}{n_l!} + \log n_l! = n_l \log n \tag{8}$$

位。

2. 我们从 $E(G)$ 中删除所有 用于编码谢尔宾斯基图子图的边的位，节省 $\frac{n_l}{(n_l-1)\ 2}$ 位。

现在，很容易给出一个算法 $P$，它给定 $E'(G)$ 后，便能构造谢尔宾斯基图子图；输入包括图的顶点数 $n$ 、 $n_l$ 的值，以及谢尔宾斯基图顶点的特定次序$i_1 i_2 \dots i_{n_l-1} i_{n_l}$，从而得到由 $n_l$ 个顶点组成的子图。

新编码的长度至多为

$$l(E'(G)) = l(E(G)) + n_l \log n - \frac{n_l(n_l-1)}{2}. \tag{9}$$

给定 $n$ 的值，程序 $P$ 可以从 $E'(G)$ 重构 $E(G)$。因此

$$C(E(T)|n, P) \leq l(E'(T)). \tag{10}$$

由于 $G$ 不可压缩，必有 $l(E'(G)) \geq l(E(G))$。由 (9) 可知，这只有在以下条件成立时才可能：

$$n_l \log n \geq \frac{n_l(n_l-1)}{2} \Leftrightarrow n \geq 2^{\frac{n_l-1}{2}} \tag{11}$$

这是所要求的。 □

定理4指出，任何含有少于 $2^{\frac{n_l-1}{2}}$ 个顶点的不可压缩图都不能包含 $S_l$作为一个诱导 子图。因此，不可压缩图在其边的单色 诱导子图中也不能包含 $S_l$ ，无论其边采用任何二染色。于是，由推论1和定理4，我们得到如下结论：

**定理5**。 除 $l$的有限个取值外，任何具有 $r_{ind}(S_l)$ 个顶点的不可压缩图都不能将 $S_l$作为诱导子图。

证明。 设 $G$ 是一个具有 $n$ 个顶点且包含 $S_l$作为诱导子图的不可压缩图。由推论1可知， $r_{\text{ind}}(S_l) \leq n_l^{c_d}$，因此 $n \leq n_l^{c_d}$。此外，根据定理4，界 $n \geq 2^{\frac{n_l-1}{2}}$ 成立。因此， $n \leq n_l^{c_d}$仅对 $n_l$的有限个取值成立，其基数取决于常数 $c_d$；这是因为 $n$的增长速度相对于 $n_l$ 呈指数级，而 $r_{\text{ind}}(S_l)$的界 $n_l^{c_d}$相对于 $n_l$呈多项式级增长。□

此外，根据引理1，我们得到如下比定理5更强的结果：

**定理6**。 随着 $l$ 增长，几乎所有具有 $r_{ind}(S_l)$ 个顶点的图（其中的分数为$(1 - \frac{1}{r_{ind}(S_l)})$ ）都满足：对其边进行任何二着色，都不会包含 $S_l$的诱导单色副本。

证明。 按照定理4中相同的证明思路，我们现在从一个标记图 $G$ 开始，使得

$$C(E(G)|n, P) \geq \frac{n(n-1)}{2} - \log n. \tag{12}$$

根据引理 1，这些图至少占所有具有$n$ 个顶点的标记图的分数$(1 - \frac{1}{n})$。对于 $n = r_{\text{ind}}(S_l)$，这些图至少占所有具有 $r_{\text{ind}}(S_l)$ 个顶点的图的比例$(1 - \frac{1}{r_{\text{ind}}(S_l)})$。以下陈述适用于这些图全部；当 $l$ 趋于无穷时，这些图包括具有 $r_{\text{ind}}(S_l)$ 个顶点的几乎所有 可能图。

证明的其余部分与定理4的证明基本相同，令$n = r_{\text{ind}}(S_l)$，但现在必须满足以下不等式，而不是不等式 (11)：

$$n_l \log n + \log n \geq \frac{n_l(n_l-1)}{2} \Leftrightarrow n \geq 2^{\frac{n_l(n_l-1)}{2(n_l+1)}} \Leftrightarrow r_{\text{ind}}(S_l) \geq 2^{\frac{n_l(n_l-1)}{2(n_l+1)}}. \tag{13}$$

然而，由于 $r_{\text{ind}}(S_l) \leq n_l^{c_d}$对于某个常数 $c_d$ ，该常数仅取决于图顶点的最大度 $d$ ，根据定理3，

不等式 (13) 将要求 $n_l^{cd} \geq 2^{\frac{n_l(n_l-1)}{2}}$，但从 $l$ 的某个值开始，这将无法成立。□

**定义 8（规模可构造图）**。设 $\mathcal{F}$ 是一个图族。如果 $\mathcal{F}$ 中的每个图都可以由一个算法 $P_{\mathcal{F}}$ 唯一构造出来，我们就称 $\mathcal{F}$ 为规模可构造图。该算法以图的规模（即图的顶点数）以及可能提供顶点某种排序信息的一个排列作为输入。

定义8的直接推论是，谢尔宾斯基图族是规模可构造的。这是因为，一个有 $n_l$ 个顶点的谢尔宾斯基图 $S_l$，只需 $n_l$ 信息以及表示其顶点排序的置换即可描述，而我们在定理4的证明中所描述的重构算法 $P$ 正是利用该置换来重构图。

然后，定理4可以概括如下：

**定理7**。设 $\mathcal{F}$ 为一族规模可构造图。设设 $G$ 为一个带标记的不可压缩图，包含 $n$ 个顶点，且其诱导子图中包含 $\mathcal{F}$ 中的一个图，该图有 $k$ 个顶点。那么，如果 $P_{\mathcal{F}}$ 需要关于顶点排序的信息，则 $n \geq 2^{\frac{k-1}{2}}$；如果 $P_{\mathcal{F}}$ 不需要这些信息，则(1)。

证明。如果算法 $P_{\mathcal{F}}$ 除了需要 $k$ 之外，还需要顶点的排序，那么图 $\mathcal{F}$ 可以使用 $k\log n + \log k!$ 位来描述，与定理 4 中的谢尔宾斯基图的情况相同，因此该定理的证明也适用于 $\mathcal{F}$。

如果 $P_{\mathcal{F}}$ 不需要排序信息，那么我们至多需要，而不是需要定理4的证明中公式（8）所给出的比特数

$$\log \frac{n}{k!} = k\log n - \log k!. \tag{14}$$

位，因为不需要排序信息，即 $\log k!$ 位。

因此，对于不可压缩图$G$，下列不等式必须成立，正如定理 4 证明中谢尔宾斯基图的不等式 (11) 所示：

$$k\log n - \log k! \geq \frac{k(k-1)}{2}. \tag{15}$$

使用斯特林近似 $k! \approx (\frac{k}{e})^k\sqrt{2\pi k}$ 并求解 $n$，不等式 (15) 得出

$$n \geq k2^{\frac{k}{2}}\left(\frac{1}{e\sqrt{2}} - o(1)\right)$$

这是所要求的。 □

在我们的语境中，也就是以谢尔宾斯基图的形式出现一个组织有序、紧密连接的社群，上述结果具有一些有趣的推论。首先，定理4和定理7（定理4的推广）表明，要在不可压缩的、即类随机的交互主体网络中存在或有目的地形成谢尔宾斯基图这类组织结构，网络的规模必须相对于组织结构的规模呈指数级增长。较小的互连网络，即相对于谢尔宾斯基图的规模呈多项式级增长的网络，几乎可以肯定不包含此类组织结构。

## 5 不可压缩图的诱导 Ramsey 数

本节研究柯尔莫哥洛夫随机图背景下的 Ramsey 数。更具体地说，我们证明柯尔莫哥洛夫随机图规模的界，使其包含谢尔宾斯基图（或其他规模可构造图）作为子图。

**定义9**。（不可压缩图的诱导拉姆齐数）不可压缩图的诱导拉姆齐数，$r^{INC,\ \delta(n)\ ind}(H)$，定义为满足以下条件的最小整数：存在一个柯尔莫哥洛夫随机图$G$，其随机性缺陷至少为 $\delta(n)$，具有 $r^{INC,\ \delta(n)\ ind}(H)$ 个顶点，使得 $E(G)$ 的每种二染色都包含 $H$的一个诱导单色副本。

**定理8**。设 $H$ 是一个具有 $k$ 个顶点且最大度不超过 $d$的规模可构造图，其描述需要顶点排序信息（对于不需要此类信息的规模可构造图，也有类似结果）。设 $n_1 = 2^{\frac{k-1}{2}} - 1$ 和 $n_2 = r_{ind}(H)$。则对于 $\delta(n) \geq n_1 n_2 + \binom{n_2}{2}$，有：*(i)* $r^{INC,\ \delta(n)\ ind}(H)$ 存在，且*(ii)* $r^{INC,\ \delta(n)\ ind}(H) < 2^{\frac{k-1}{2}} + r_{ind}(H)$。

证明。设 $G_1$ 是一个具有 $n_1 = 2^{\frac{k-1}{2}} - 1$ 个顶点的不可压缩图，因此（根据定理7），它不包含 $H$ 作为诱导子图。由于 $G_1$ 是不可压缩的，因此有 $C(E(G_1)) \geq \binom{n_1}{2}$。

另外，令 $G_2$ 为一个有 $n_2 = r_{\text{ind}}(H) \leq k^{c_d}$ 个顶点的任意图，使得无论如何对其边进行二着色，它都包含一个 $H$ 的诱导单色副本（定理3保证了此类图的存在）。

令 $n = n_1 + n_2$。我们关注有 $G$ 个顶点的图 $n$，它们只由一个 $G_1$ 副本和一个 $G_2$副本组成，也就是说，包含两个分别与 $G_1$ 和 $G_2$同构的子图。除这两个子图中的边外，不存在其他边。

显然， $E(G)$ 的每一种二着色都包含 $H$ 的一个诱导单色副本，因为 $G_2$ 确实如此。这证明，对于 Kolmogorov 随机图，即不可压缩图（至少对于下文将定义的亏损函数 $\delta(n)$），$r^{\mathrm{INC}}{}_{,\delta(n)\ \mathrm{ind}}(H)$ 存在。

此外， $G$ 有 $n = n_1 + n_2 = 2^{\frac{k-1}{2}} - 1 + r_{\mathrm{ind}}(H)$个顶点。我们将证明，对于适当的随机性缺陷函数 $\delta(n) = \delta(n_1, n_2)$，以下成立，即 $G$ 是 $\delta(n_1, n_2)$-不可压缩的：

$$C(E(G)) \geq \binom{n_1 + n_2}{2} - \delta(n_1, n_2)$$
$$\text{with } \delta(n_1, n_2) \geq \binom{n_2}{2} + n_1 n_2. \tag{16}$$

反设，对于一个满足不等式 (16) 的随机性缺陷函数 $\delta(n_1, n_2)$，有 $C(E(G)) < \binom{n_1+n_2}{2} - \delta(n_1, n_2)$。则下述结论成立：

$$C(E(G)) < \binom{n_1 + n_2}{2} - \delta(n_1, n_2) \leq \binom{n_1 + n_2}{2} - \left[\binom{n_2}{2} + n_1 n_2\right] = \binom{n_1}{2}. \tag{17}$$

我们将描述一种算法，可以根据对 $G$ 的 $C(E(G)) < \binom{n_1}{2}$ 位描述重构 $G_1$，这与我们的假设相矛盾：$C(E(G_1)) \geq \binom{n_1}{2}$，即 $G_1$ 是不可压缩的。

假设我们有对 $G$ 的 $C(E(G)) < \binom{n_1}{2}$ 位描述。那么我们可以如下重构 $G_1$。首先，我们使用其 $C(E(G))$ 位描述重构 $G$。然后，我们通过例如 深度优先搜索（*DFS*）算法（见 [14]）计算它的连通分量。设这些连通分量为 $G_{i_1}, G_{i_2}, \ldots, G_{i_s}$。其中一个连通分量必定是图 $G_2$。其余连通分量的并集应当是图 $G_1$。然而，如果 $G_1$ 是一个连通图，那么除了 $G_2$之外，可能只存在一个连通分量。下一步是确定 $G_2$。如果我们成功完成这项任务，就确定了 $G_1$：它是由 $G$ 中不属于 $G_2$的顶点和边构成的图。

为了识别 $G_2$ 我们逐一处理连通分量 $G_{i_1}, G_{i_2}, \ldots\ldots, G_{i_s}$。我们在这里应当谨慎，因为 $G_2$ 可能不是一个连通的图，因此我们需要找到全部 其连通分量。关键在于

观察到， $G_2$ 不能包含这样的连通分量：其边的任何二染色都不含有 $H$ 的诱导单色副本。否则，我们可以去掉该连通分量，使一个更小的图满足 $r_{\text{ind}}(H)$ 的定义，这就与其极小性要求矛盾。

设 $G_{i_l}$ 为当前检查的连通分量。我们生成该连通分量的所有可能的边的二染色，并且对于其中的每一个，检查 $G_2$ 是否包含 $H$ 的单色副本。注意到这只有对 $G_2$ 连通分量才可能成立，因为所有其他连通分量都属于 $G_1$，后者是一个顶点数少于 $2^{\frac{k-1}{2}}$ 的不可压缩图，因此，根据定理4，其边的二染色（以及相应地，其连通分量的二染色）不可能包含 $H$ 的单色副本。因此，在确定了 $G_2$ 后，我们可以将 $G_1$ 重构为 $G$ 中包含其余连通分量的子图。这样，我们成功地使用少于 $\binom{n_1}{2}$ 位重构了 $G_1$，这就产生了矛盾。因此，$C(E(G)) \geq \binom{n_1+n_2}{2} - \delta(n_1, n_2)$，也就是说， $G$ 是 $\delta(n_1, n_2)$-不可压缩的。

总之， $G$ 是一个 $n = n_1 + n_2$ 图，对于 $\delta(n) \geq \delta(n_1, n_2)$，它是 $\delta(n)$-不可压缩的，并且其边的所有二染色都包含一个 $H$的诱导单色副本。因此可得， $r^{\text{INC}}{}_{,\delta(n)\ \text{ind}}(H) \leq n = 2^{\frac{k-1}{2}} - 1 + r_{\text{ind}}(H)$，或者用更简单的形式表示为， $r^{\text{INC}}{}_{,\delta(n)\,\text{ind}}(H) < 2^{\frac{k-1}{2}} + r_{\text{ind}}(H)$。□

由定理 4 和定理 8 可得：

**推论2**。 对于每个图 $H$ 上有 $k$ 个顶点且最大度至多为 $d$，都有 $2^{\frac{k-1}{2}} \leq r^{INC}{}_{,\delta(n)\ ind}(H) < 2^{\frac{k-1}{2}} + r_{ind}(H)$.

请注意，由于 $n = n_1 + n_2$， $G$ 的规范表示需要 $|E(G)| = \binom{n_1+n_2}{2}$ 位，而我们令 $\delta(n) = \delta(n_1, n_2) = n_1 n_2 + \binom{n_2}{2}$。因此，将 $k$ 视为变化的参数时， $\delta(n)$ 的阶为 $\sqrt{|E(G)|}$。我们无法证明对于小于此值的亏损函数 $\delta(n)$，定理8仍然成立，例如 $\delta(n) = \log|E(G)|$。不过，我们认为这是不可能的，因为当一个图具有某种规律性时，在我们的例子中即为具有 $n_l$ 个顶点的谢尔宾斯基图 $S_l$，其复杂度会降低，因为可以利用这种规律性来缩短其描述。

## 6 Conclusionsanddirections for further research

社会是由多方面、自主且相互作用的主体及其相互关系构成的复杂人类创造物。这是一个常见主题。

大量研究工作表明，在主体复杂网络中，随着（甚至是随机的）相互作用和关系不断演化，重要的现象和模式会随之涌现。这些结果依托于关于主体相互作用模型的成熟数学和物理理论，例如随机图论和复杂网络理论。

本文考察了两种成熟的数学理论框架，它们以有限对象复杂度这一概念为研究目标；在本文中，这些对象是社会及其交互主体：*Kolmogorov* 复杂度和拉姆齐理论。这两种理论各自从不同视角研究在大型、不断演变的结构（例如交互主体构成的社会网络）中，某些子结构出现（或不出现）的条件，并由此估算这些结构需要达到多大规模，才能包含此类规律性。我们运用这两种理论中的一些内容，研究不断演变的社会中具有某些理想属性的规则结构或模式的出现情况。我们关注的属性之一是紧密连接性，换言之，这一属性描述的是由相互沟通的主体构成的结构，其中的成员彼此紧密互动，并坚决维护其群体的连贯性和观点，以对抗其所处的更大结构。

具体而言，借助定理5，我们证明了谢尔宾斯基图作为一种高度组织化的结构，只有当不断演变的图（社会）达到足够规模时才会出现；该规模相对于组织化结构的大小呈指数级增长。借助定理8和推论2，我们给出了图（社会）规模的界限，确保其中包含任意度有界的组织化结构（即顶点之间的关系）；这些顶点即社会成员。

我们希望，本研究能够推动人们进一步运用有关复杂结构及其长期演化特性的丰富数学理论。通过这种方式，我们认为可以加强运用源自精确科学与形式系统的方法研究社会科学的努力。

## 参考文献

1. J.Beck，关于路径、树和回路的规模Ramsey 数II，载于：*Mathematics of Ramsey theory* 5，1990，34 45，《算法与组合学》，施普林格，柏林，–。2.B.Bollob´as，第二版。剑桥大学出版社，2001 年。3. G.H. Chaitin，“关于计算有限二进制序列的程序长度”，*Journal of the ACM* 1k，第 547–570 页，1966 年。

4. V. Chv´atal、V. R¨odl、E. Szemer´edi 和 W.T.Trotter Jr.有界最大度图的 Ramsey 数。 *J. Combin. Theory Ser. B* **34** (1983), 239–243。5.D. Conlon、J. Fox 和 B. Sudakov。 *On two problems in graph Ramsey theory* *Combinatorica* **32** 2012 513 535 (), –。6. R.Diestel, *Graph theory*, 第2版, 施普林格, 1997 年。7. P.E. Haxell、Y. Kohayakawa 和 T. ´Luczak, 圈的诱导 size-Ramsey 数, *Combin. Probab. Comput.* **4** (1995), 217–240。8.A.N.Kolmogorov。 *Three approaches to the quantitative definition of information*。《信息传输问题》, 1(1):1–7, 1965 年。9.M.Li 和P.M.B. Vit ´ *An Introduction to Kolmogorov Complexity and its Applications* 4 2019 anyi。。施普林格, 纽约, 第版, 。10.R.J.Solomonoff, "归纳推理形式理论" *Information and Control* 7 1, 1 22, 1964 , 第卷/第页 –。11.T. ´Luczak 和 V. R¨odl, 有界最大度图的诱导 Ramsey 数, *J. Combin. Theory Ser. B* **66** (1996), 324–333。12. F.P. Ramsey, 关于数理逻辑的一个问题, *Proc. London Math. Soc.* **30** (1930), 264–286。13. H. Peyton Young。社会网络中的创新扩散。在 *Proc. The ,, Economy as a Complex Evolving System vol. III* *Lawrence E. Blume and Steven ,,*。14. R. *N. Durlauf (eds.) Oxford University Press (2003)* Sedgewick 和 K. Wayne。 *Algorithms*。Addison-WesleyProfessional, 第 4 版, 2011 年。15.B.L. van der Waerden, Baudet 猜想的证明, *Nieuw. Arch. Wisk.* ( ), –。**15** 1927 212 216