|  |  |
| --- | --- |
| 来源 | 2015 HPDC Understanding Graph Computation Behavior to Enable Robust Benchmarking |
| 摘要 | Graph processing is widely recognized as important for a growing range of applications, including social network analysis, machine learning, data mining, and web search. Recently, many performance assessments and comparative studies of graph computation have been published, all of which employ highly varied ensembles of algorithms and graphs. To explore the robustness of these studies, we characterize how behavior varies across a variety of graph algorithms on a diverse collection of graphs (size and degree distribution). Our results show that graph computation behaviors form a very broad space, and inefficient exploration of this space will possibly lead to an ad-hoc study as well as a narrow understanding of graph processing performance. Hence, we consider constructing an ensemble of experiments, e.g. a benchmark suite, to most effectively explore graph computation behavior space. We study different ensembles of graph computations, and define two metrics, spread and coverage, to quantify how efficiently and completely an ensemble explores the space. Our results show that: (1) an ensemble drawn from a single algorithms or a single graph may unfairly characterize a graphprocessing system, (2) an ensemble exploring both algorithm diversity and graph diversity improves the quality significantly, but must be carefully chosen, (3) some algorithms are more useful than others for exploring the space, and (4) it is possible to reduce benchmarking complexity (i.e. number of algorithms, graphs, etc.) while conserving the benchmarking quality.  图计算被广泛运用在社交网络分析、机器学习、数据挖掘和网页搜索中，变得越来越重要。最近，许多图计算相关的测评和比较相关的研究已经被发表，所有这些都采用高度多样化的算法和图形集合。为了探索这些研究的鲁棒性，我们描述了行为在不同图形集合上的各种图形算法（尺寸和度数分布）是如何变化的。我们的研究结果表明图计算的表现行为具有非常广阔的空间，在这个空间里，低效的探索方式将会使得研究变成一个ad-hoc的研究，而且会对图计算性能产生狭隘的理解。  因此，我们考虑构建一组实验，例如 一个基准套件，以最有效地探索图计算行为空间。 我们研究图计算的不同集合，并定义两个度量，扩展（spread）和覆盖（coverage），以量化集合探究空间的有效性和完整性。 我们的结果表明：（1）从单个算法或单个图形绘制的集合可能不公平地表征图形处理系统，（2）集合探索算法多样性和图形多样性显着提高质量，但必须仔细选择， 3）一些算法比其他算法更有用于探索空间，以及（4）可以在保持基准质量的同时降低基准复杂度（即算法，图形等的数量）。 |
| 对比目的 |  |
| 数据集 |  |
| 应用 |  |
| 测试方法 |  |
| 研究问题 |  |
| Ensemble算法发现 |  |
| Mahout算法发现 |  |
| 结论 |  |
| 相关有趣的参考文献 |  |
| 启发 |  |

# Understanding Graph Computation Behavior to Enable Robust Benchmarking

**Abstract**

图计算被广泛运用在社交网络分析、机器学习、数据挖掘和网页搜索中，变得越来越重要。最近，许多图计算相关的测评和比较相关的研究已经被发表，所有这些都采用高度多样化的算法和图形集合。为了探索这些研究的鲁棒性，我们描述了行为在不同图形集合上的各种图形算法（尺寸和度数分布）是如何变化的。我们的研究结果表明图计算的表现行为具有非常广阔的空间，在这个空间里，低效的探索方式将会使得研究变成一个ad-hoc的研究，而且会对图计算性能产生狭隘的理解。

因此，我们考虑构建一组实验，例如 一个基准套件，以最有效地探索图计算行为空间。 我们研究图计算的不同集合，并定义两个度量，**扩展（spread）**和**覆盖（coverage）**，以量化集合探究空间的有效性和完整性。 我们的结果表明：（1）从单个算法或单个图形绘制的集合可能不公平地表征图形处理系统，（2）集合探索算法多样性和图形多样性显着提高质量，但必须仔细选择， 3）一些算法比其他算法更有用于探索空间，以及（4）可以在保持基准质量的同时降低基准复杂度（即算法，图形等的数量）。

**1. Introduction**

万维网和社交网络的快速增长产生了海量的图数据，这些图数据广泛应用在社交网络，推荐系统，交互网络，web图等的建模分析中。来自这些真实世界中的图非常巨大，再加上运行在这些数据上的复杂的算法：他们迫切需要高效可扩展的图形处理系统。

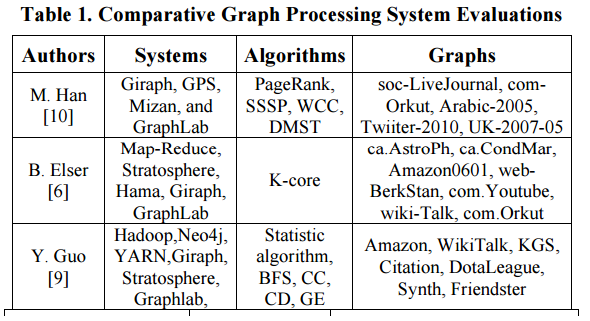
为了满足这种需要，研究界已经创建了许多图形处理系统，包括Pregel [19]，Giraph [1]，GraphLab [7]，SNAP [24]，TurboGraph [11]，Mizan [16] ，和GraphChi [17]，旨在满足图形处理的挑战：（i）极端规模（100亿的边缘），（ii）不规则的计算结构（难以聚合多个顶点操作），（iii）局部性差， （iv）平行度的宽泛变化。

[图的结构是变化多样的]

混淆因素是图形和图形算法的性质的广泛变化。 图形特征，包括图形大小，顶点度和顶点/边缘权重，很大程度上取决于应用程序域。 例如，来自Web数据共享[27]的最大图形具有35亿个顶点（网页）和128亿个边缘（超级链接）。 另一个网络图形，“维基百科投票网络”[24]，只包含7,115个顶点和103,689个边。 顶点度也因域而异。 在社交网络中，流行的Twitter或Facebook帐户拥有数百万的关注者，而其他人只有少数。 但是在从线性求解器导出的图中，顶点具有低的几乎均匀的程度。

[算法的特点也是迥异的]

图算法的动态特性（**计算强度**，**通信强度**和**顶点活动分布的变化**）也表现出很大的多样性。 例如，顶点处的计算可以基于算法而广泛变化。 通信强度通常取决于顶点度和算法。 顶点的活动分数是算法的强函数 - 在PageRank中，所有顶点开始活动，分数逐渐减少，而在单源最短路径（SSSP）中，只有源顶点开始活动，但活动分数快速增长。



我们的目标是提供对图算法和结构对图形处理系统的性能影响的更系统的理解，从而实现鲁棒的系统评估。

为了理解和量化图计算行为的多样性，我们从不同应用领域中选择一部分算法来研究其动态属性。 使用GraphLab作为实验载体，算法在不同大小和结构的合成图的集合上运行。 为了表征行为，我们测量从**活动顶点**到**计算强度**的关键属性，表征图算法和结构的行为变化。不仅算法的行为不同，而且一些图算法非常依赖输入图的特征。 这与许多数值，科学并行计算显着不同。 事实上，我们的行为度量：**活动分数**，**计算强度**和**通信强度**显示对图结构的高敏感性。

基准化包括了解行为空间的性能。 因此，为了系统地理解图形处理行为，我们使用也影响性能的基本图形计算属性来定义向量空间。 为了分析运行集合的属性，我们定义集合度量：**扩展**和**覆盖**，捕获一组图形算法对探索行为空间的程度。 使用这些指标，我们评估不同运行集合的基准质量。 我们的研究结果表明，无论是单一的算法与各种图形结构的组合，还是多个算法与单个图形的组合，都无法很好的探索行为空间，这引起图形处理系统的反复评价。

为了达到最佳的传播和覆盖，我们还搜索多个算法和多个图组合的运行情况。 **我们的研究表明，一小组精心挑选的集合可以更有效地对行为空间进行抽样，这提供了构建高质量基准套件的机会。** 然后在某些约束下，我们系统设计了几个组合，并且对他们进行了评估。**本文的贡献如下**：

（1）我们进行了一系列实验不同的算法和图形，证明了图形计算行为的五个维度的1000倍变化。

（2）我们定义两个集合度量，扩展和覆盖，以衡量实验集合如何有效和彻底地探索图计算行为空间。

（3）我们证明从单个算法或单个图形绘制的集合是一个差的基准集合，并且集合探索算法多样性和图形多样性提供200％更好的传播和30％更好的覆盖。

（4）我们对具有高扩展和覆盖的最佳集合的洞察表明，一些算法，包括K均值，交替最小二乘和三角计数，在行为空间探索中比其他算法更有用。

（5）我们发现，算法分集的仔细减少最小化了扩展和覆盖的损失，进一步优化可能通过减少运行时间。

**2. Graph Algorithms and Structure**

我们描述了各种图形计算问题，以及在其中出现的算法和图形结构。 我们从不同的应用领域中选择了他们广泛变化的行为的算法。

**2.1 Graph Algorithms**

（1）**Graph Analytics**(GA) 图分析专注于数据挖掘，特别是大图中的相关关系。我们选择了六个算法：连通子图（CC）,K-Core分解（KC），三角形计数（TC），单源点最短路径（SSSP），PageRank（PR）和近似半径（AD）。

（2）**Clustering**聚类算法是根据对象的相似性对对象进行分类。我们选择K-Means（KM）算法将n个顶点分成k个簇。

（3）**Collaborative Filtering** (CF)协同过滤是推荐系统用来预测用户给予项目（例如电影或书籍）的“评级”或“偏好”的技术。CF的一般方法是通过矩阵分解来学习所有的用户因子向量和项目因子向量。再此我们选择了四个算法：ALS, NMF, SGD, SVD.

（4）**Other Algorithms** 包括Jacobi，LBP和DD.

**2.2 Graph Structure**

图形功能包括**大小**，**顶点值**，**边权重**和**连接性**（度分布）。所有这些功能在应用程序域之间显着不同。 例如，GA应用程序使用图形作为真实世界图形的抽象; 聚类应用程序使用图形来表示数据点列表; CF应用程序使用图形来表示推荐矩阵; 线性求解器使用图来表示矩阵。

**3. Methods**

在这一章节，我们介绍用于实验的系统配置，每个域的数据集，以及定义我们实验中使用的所有性能指标。