

假设分区完毕,调整 $k - 1$ ,

对于超边 $E$ ,其当前 $k - 1$ 计为 $k_{old}$ ,目的将其 $k - 1$ 降低为 $k_{old} - 1$ ,需要变更 $\frac{E.degree}{k_{old}}$ 个点计为 $n_i$ ,一个点的平均度数为 $D(n_i)$

大小为 $x$ 的超边随机分到 $p$ 个分区, $Ex(x, p)$ 表示其跨分区数 $k - 1$ 的期望,易得:

$$Ex(x, p) = p[1 - (1 - \frac{1}{p})^x]$$

假设 $x$ 的概率密度函数 $f(x)$ 服从某分布(此处不妨假设服从幂率分布) $f(x) \sim \frac{1}{x^\alpha}$ ,对于某分区其超边表

目前现象:

## 现象一

现象:对于同一数据集,当分区数 $p$ 增大时,参数sheild需要变大,且运行时间增大

分析:对于同一数据集,当分区数较小时大部分小度数超边可以在某一个或较少的几个分区内构成较为密集的团,从结果而言此时对于度数较大的超边,很有可能会因此跨过较多分区(甚至跨过所有分区),此时在运行过程中对大度数的超边更新其对应节点的值会带来负面的影响(更新权值的目的在于提高该条超边所包含节点的权值,减少该超边跨分区数,但如果从结果出发这样的操作如果不能(或者只能减少一点)分区数,那么此时这样的权值更新操作反而会使得部分小度数节点原先能构成密集子图的点权值被稀释了,进而导致分区质量下降),因此对于分区数较少时,应当设置较小的sheild值,牺牲更多的大度数超边的分区质量,提高大多数小度数超边的分区结果

另一方面,若分区数增大时,即使小度数节点也可能不得不跨过很多分区,此时大度数定点的稀释作用就减少了,此时提高sheild阈值可以在不提升小度数超边跨分区数的同时,兼顾更多的大度数定点使得最后的分区质量更优(定性分析),但是如果考虑大度数定点则必然会带来运行时间的上升(见现象二)

## 思路

问题的实质在于,对于不同规模的超边其跨分区彼此之间的影响,当分区数小时,应当更关注小度数超边的分区情况,对于分区数较大时适当考虑较大超边的收益,对于该情况我们进行如下建模:

假设给定超图 $G$ (假设超图的超边之间独立),其超边度数服从某一分布(不妨设幂率分布)即 $f(|e_i|) \sim \frac{1}{|e_i|^\alpha}$ ,根据某一算法(不妨假设随机分区)使得经过该分区算法之后对于超边 $e_x$ 其当前 $cross(e_x)$ 计为 $k_{old}$ ,若试图将 $cross(e_x)$ 减小为 $k_{new}$ ,则需要对当前分区结果进行微调,调整步骤如下:

- 不妨设 $k_{new} = k_{old} - 1$ ,调整后相当于将原先 $k_{old}$ 个分区中选择一个分区的点,在尽可能不增加其他超边跨分区数的情况下,划分到剩下 $k_{new}$ 个分区中,因此需要调整的点数期望为 $\frac{|e_x|}{k_{old}}$
- 调整可以视为两步,第一步将 $\frac{|e_x|}{k_{old}}$ 点从某分区中移除,这一步可能伴随该分区内超边表的变化(减少),其减少的期望为:

对于随机分区算法任意超边,设某分区内该超边涉及顶点数的 $R(e_i)$ (其期望可以表示为 $\frac{|e_i|}{p}$ ),而该超边其该分区的跨分区数指标的贡献以 $(1 - \frac{1}{p})^{|e_i|}$ 的概率为1,其他情况下为0.因此该分区的超边数期望可以表示为 $\sum_{e_i \in E} (1 - \frac{1}{p})^{|e_i|}$ ,设该分区中点数为 $n_p$ ,从该分区中移除 $x$ 个点,使得该分区中超边 $e_i$ 的贡献从1变为0的概率为 $\frac{C_{n_p}^{x-R(e_i)}}{C_{n_p}^x}$ ,因此对于该分区,其超边表大小(减少)变化的期望为:

$$Remove(x) = \sum_{e_i \in E} \sum_{j=1}^{|e_i|} C_{|e_i|}^j * (1 - \frac{1}{p})^{|e_i|-j} * (\frac{1}{p})^j * \frac{C_{n_p}^{x-j}}{C_{n_p}^x}$$

上式可以理解为:

$$\sum_{e_i \in \text{所有超边}} \text{该分区中含有 } e_i \text{ 的概率} * \frac{\text{将 } e_i \text{ 完全移除的方案数}}{\text{总方案数}}$$

- 将这 $\frac{|e_x|}{k_{old}}$ 个点重新分配给其他 $k_{new}$ 个分区该操作会使得这 $k_{new}$ 个分区的超边数增加,增加的期望为:

一个点所涉及的超边数为 $R(n_i)$ (其期望可以表示为 $\sum_{e_i \in E} \frac{|e_i|}{|V|}$ ),而一个顶点加入到一个分区可能带来的分区超边数增长可以表示为:

$$\sum_{e_i \in E} \frac{|e_i|}{|V|} * (1 - \frac{1}{p})^{|e_i|-1}$$

同时我们注意到其中至少对于目标期望减少的超边在原分区中已经出现因此上式进行修正后需要减去 $\frac{|e_x|}{|V|} * (1 - \frac{1}{p})^{|e_x|-1}$

上式可以理解为:

$$\sum_{e_i \in \text{所有超边}} e_i \text{ 中包含 } n_i \text{ 的概率} * \text{当前分区中不含有 } e_i \text{ 的概率}$$

