

高等教育学业表现同群效应研究

——基于校园卡大数据

邱磊菊 李 寅 段月姣

目录

附录 I 稳健性检验	1
附录 II 情感投入机制的进一步分析	3
附录 III 附图及附表	5
参考文献	7

附录 I 稳健性检验

首先,为避免阈值设定存在一定的任意性问题,导致的朋友识别低估的情形,稳健性检验中尝试了不同阈值设定的检验。保持 5 分钟的时间窗口不变的条件下,依次将阈值设定为 10 次、20 次、30 次、40 次、50 次,并根据结果分别构建社交网络,并进行检验。表 I 1 分别显示了阈值为 10 次、20 次、30 次、40 次、50 次以及原文中 28 次的对比结果。结果显示,学业表现系数显著为正,并且系数值在 0.2-0.3 左右,未产生巨大变化,表明该结果对阈值次数的设定不是很敏感,结果较为稳健。

表 I 1 不同阈值设定下的学业表现同群效应检验

	(1)	(2)	(3)	(4)	(5)	(6)
变量名	社交网络主体学生学业表现: <i>Score</i>					
	阈值=10	阈值=20	阈值=30	阈值=40	阈值=50	阈值=28
<i>GP_Score</i>	0.234*** (0.061)	0.361*** (0.032)	0.298*** (0.024)	0.222*** (0.020)	0.200*** (0.020)	0.304*** (0.025)
<i>Constant</i>	1.759*** (0.143)	1.477*** (0.074)	1.630*** (0.055)	1.815*** (0.048)	1.863*** (0.047)	1.613*** (0.059)
控制主体特征变量	YES	YES	YES	YES	YES	YES
控制朋友特征变量	YES	YES	YES	YES	YES	YES
固定效应	YES	YES	YES	YES	YES	YES
N	7578	6866	5825	4707	3680	6160
R ²	0.102	0.110	0.119	0.119	0.126	0.115

其次,同一专业的同学选课较为集中,同一时间点下课可能导致学生刷卡时间较为接近。为尽量解决这一因素产生的结果偏误,文章剔除了在午饭前与晚饭前有课的学生去食堂的刷卡记录,重新构建网络进行检验。基于样本学校课表安排:上午 1-4 节课,下午 5-8 节课,晚上 9-10 节课。第 4 节 12:00 下课,第 8 节 17:30 下课。筛选出第 4 节课和第 8 节课有课的学生,相应剔除他们 12:00~12:30 以及 17:30~18:00 的就餐刷卡记录,进一步排除因为同一时间下课导致的刷卡时间接近问题。删除该部分刷卡记录后,按原文的方法重新构建社交网络,得到回归数据集,样本从原来的 6160 变为 4432,回归结果如表 I 2 所示。第 (3) 列影响系数为 0.252,相对小于原有结果,但在 1% 显著性水平上显著。结果相对稳健,影响效力在 0.3 左右。

表 I 2 剔除饭前有课学生就餐的刷卡记录后重建社交网络的对比分析

	(1)	(2)	(3)	(4)	(5)	(6)
变量名	剔除后学生社交网络学业表现: <i>Score</i>			全样本学生社交网络学业表现: <i>Score</i>		
<i>GP_Score</i>	0.311*** (0.020)	0.242*** (0.020)	0.252*** (0.020)	0.387*** (0.024)	0.298*** (0.024)	0.304*** (0.025)

<i>Constant</i>	1.596*** (0.046)	1.758*** (0.046)	1.741*** (0.048)	1.415*** (0.056)	1.627*** (0.056)	1.613*** (0.059)
控制主体特征变量	NO	YES	YES	NO	YES	YES
控制朋友特征变量	NO	NO	YES	NO	NO	YES
固定效应	NO	YES	YES	NO	YES	YES
N	4432	4432	4432	6160	6160	6160
R ²	0.053	0.126	0.129	0.040	0.115	0.115

第三，作者进一步考虑周末在学校就餐的刷卡数据。周末与工作日食堂就餐的比例呈现“二八”关系，按照样本年份统计周末在食堂就餐占全部时间就餐比例，分别为 20.58%、21.26%、20.20%。因而，文章使用周末就餐的学生作为子样本，重新识别社交网络进行稳健性检验。具体汇报在表 I 3 中。周末就餐群体中的同群效应在学业表现中系数为 0.249，相对小于工作日就餐的社交群体的影响效应系数（0.338）。

表 I 3 周末与工作日就餐学生社交网络对比分析

变量名	(1)	(2)	(3)	(4)	(5)	(6)
	周末就餐社交网络学业表现： Score			工作日就餐社交网络学业表现： Score		
<i>GP_Score</i>	0.305*** (0.028)	0.242*** (0.027)	0.249*** (0.028)	0.449*** (0.031)	0.339*** (0.030)	0.338*** (0.032)
<i>Constant</i>	1.606*** (0.065)	1.761*** (0.063)	1.756*** (0.066)	1.270*** (0.071)	1.534*** (0.070)	1.533*** (0.074)
控制主体特征变量	NO	YES	YES	NO	YES	YES
控制朋友特征变量	NO	NO	YES	NO	NO	YES
固定效应	NO	YES	YES	NO	YES	YES
N	6490	6490	6490	6778	6778	6778
R ²	0.018	0.104	0.109	0.031	0.108	0.110

附录 II 情感投入机制的进一步分析

对学业的情感投入主要涉及学习满足感等心理活动,通常通过评估学生对学校和教师的情感反应来实现 (Schnitzler 等, 2021)。教育部中国高等教育学会高等财经教育分会的大学生学情调查中, 含有“你对所学专业的满意程度”“重新选择时你仍会选择本专业的可能性”对所学专业满意度的问题, 以及“你对学校的满意程度”“重新选择时你仍会选择本校的可能性”对学校满意度的调查, 在一定程度上反应了学生在群体中的情感感知, 因此本文根据上述问题构建相关情感变量, 分别记 *S-major*、*S-school*。具体来说, 当学生对一个问题给予正面评价时, 变量取值为 1, 如果都是满意评价时, 取值为 2, 否则为 0。类似的, 本文构建学生对于学校的情感变量 *S-school*。将朋友成绩对上述情感投入代理变量进行回归检验, 结果汇报于表 II 1 Panel A、Panel B 中第 (1)、(3)、(5) 列, 结果均显著为正, 说明朋友成绩对学生的情感投入存在积极正向显著影响, 偶数列作为上述情感投入代理变量对学生成绩的回归结果, 系数均显著为正, 说明学生对所学专业和学校越满意, 成绩就越好。特别地, 对于成绩相对较低的学生, 朋友的成绩对其在专业满意度上起到更为显著的影响, 而成绩较高的学生, 朋友的成绩对其在学校满意度上起到了更大的效用。以上检验证实了假设二, 即朋友学业表现越好, 主体学生内在情感投入越积极, 学业表现越好。

表 II 1 同群效应的情感投入机制的深入分析

Panel A	(1)	(2)	(3)	(4)	(5)	(6)
情感投入的 渠道--专业 满意	全样本		较低成绩群体		较高成绩群体	
变量名	<i>S-major</i>	<i>Score</i>	<i>S-major</i>	<i>Score</i>	<i>S-major</i>	<i>Score</i>
<i>GP_Score</i>	1.199*** (0.354)		0.783* (0.429)		0.992 (0.625)	
<i>S-major</i>		0.009*** (0.001)		0.004*** (0.001)		0.002*** (0.001)
<i>Constant</i>	-1.983** (0.835)	2.319*** (0.006)	-0.941 (1.015)	2.282*** (0.008)	-1.594 (1.464)	2.357*** (0.003)
控制主体特征变量	YES	YES	YES	YES	YES	YES
控制朋友特征变量	YES	YES	YES	YES	YES	YES
固定效应	YES	YES	YES	YES	YES	YES
N	6160	6160	3080	3080	3080	3080
R ²	0.040	0.108	0.047	0.086	0.059	0.052

Panel B	(1)	(2)	(3)	(4)	(5)	(6)
情感投入的 渠道--学校 满意	全样本		较低成绩群体		较高成绩群体	

变量名	<i>S-school</i>	<i>Score</i>	<i>S-school</i>	<i>Score</i>	<i>S-school</i>	<i>Score</i>
<i>GP_Score</i>	1.001*** (0.359)		0.871** (0.444)		1.096* (0.628)	
<i>S-school</i>		0.003*** (0.001)		0.002 (0.001)		0.001** (0.001)
<i>Constant</i>	-1.503* (0.845)	2.324*** (0.006)	-1.162 (1.051)	2.284*** (0.008)	-1.769 (1.471)	2.357*** (0.003)
控制主体特征变量	YES	YES	YES	YES	YES	YES
控制朋友特征变量	YES	YES	YES	YES	YES	YES
固定效应	YES	YES	YES	YES	YES	YES
N	6160	6160	3080	3080	3080	3080
R ²	0.042	0.095	0.046	0.083	0.055	0.050

附录 III 附图及附表

（一）社交网络构建详细说明

本文设定校园卡交易 5 分钟为一个“时间窗口”（详见示意图 A1），餐厅同一楼层的所有档口为一个“统计单元”。调查显示，学生及其朋友通常选择同一楼层餐厅，可能在不同档口进行消费，前后存在时间差，大约 5 分钟，如果在同一档口消费，通常一起排队不超过 5 分钟。基于校园卡的每一条刷卡记录时间，统计此笔交易记录往后 5 分钟内的交易记录，将其记作两两“共现”一次。图示中红实线框与虚线框中的交易记录为两个相邻“时间窗口”，两个相邻“时间窗口”内重复“共现”的记录仅保留一次。因此，本文主结果中构建社交网络基于同一“统计单元”5 分钟以内，所有发生校园卡交易的学生，而在稳健性中，对时间窗口延长至 10 分钟进行再次检验。

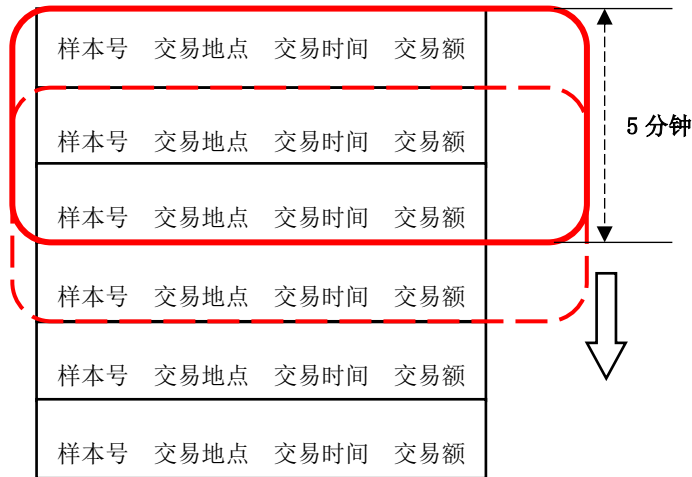


图 A1 “时间窗口”算法的示意图

基于上述算法，对校园卡交易数据进行处理，加总所有“共现”次数，得到 15,763,881 人次的“共现”记录。依据统计学“小概率事件”发生概率不超过 5%的计算法则，本文将每学期潜在朋友列表中前 5%“共现”记录保留，提取样本中五个学期“共现”频数的最大值（28），并将次数记作数据集最低阈值，从而获取“共现”频数 ≥ 28 次的记录，也即是将频数小于 28 次的记录视为陌生人巧遇，而非朋友“共现”。进而进行缺失值处理，最终保留“共现”记录仅占总记录的 0.71%，即只有 0.71%的记录被认定为朋友“共现”，余下的 99.29%的记录被判断为陌生人偶遇而予以筛除。具体记录如表 A1 所示。

表 A1 “共现”记录统计

学期	“共现”数	倒序排列前 5%对应的频数	缺失值处理后“共现”数	保留比例
一	156,351	24	2,056	1.31%
二	264,455	28 (阈值)	4,621	1.75%
三	2,750,572	19	14,072	0.51%
四	4,444,006	23	54,588	1.23%
五	8,148,497	19	36,838	0.45%
总计	15,763,881	——	112,175	0.71%

(二) 内生性检验

借鉴 Martin et al.(2020), 使用前一学期学生形成的社交网络对本学期学生学业表现进行回归, 结果如表 A2 所示, 系数依然显著。

表 A2 前一学期形成的社交网络的同群效应

变量名	(1)	(2)	(3)
	前一学期的社交网络对当期学生学业表现: <i>Score</i>		
<i>GP_Score</i>	0.293*** (0.037)	0.166*** (0.037)	0.171*** (0.039)
<i>Constant</i>	1.635*** (0.085)	1.938*** (0.086)	1.931*** (0.091)
控制主体特征变量	NO	YES	YES
控制朋友特征变量	NO	NO	YES
固定效应	NO	YES	YES
<i>N</i>	3246	3246	3246
<i>R</i> ²	0.019	0.105	0.107

参考文献

- [1] Martin, D. D., A. C. Wright, and J. M. Krieg, “Social Networks and College Performance: Evidence from Dining Data”, *Economics of Education Review*, 2020, 79, 102063.
- [2] Schnitzler, K., D. Holzberger, and T. Seidel, “All better than being disengaged: Student engagement patterns and their relations to academic self-concept and achievement”, *European Journal of Psychology of Education*, 2021, 36, 627-652.

注：该附录是期刊所发表论文的组成部分，同样视为作者公开发表的内容。如研究中使用该附录中的内容，请务必在研究成果上注明附录下载出处。