Carnegie Mellon University
Software Engineering Institute

MEMBER RETENTION DATA REPORT TO INCOSE

Sarah Sheard and Mike Konrad
December 2018

Executive Summary

Sarah Sheard, of Carnegie Mellon University's Software Engineering Institute, performed the analyses and Mike Konrad, also at CMU SEI, consulted and reviewed.

Member data provided to us was received, cleaned, redacted as to any data not needed for the analyses, and then manipulated for numerical analysis first and causal analysis second. Causal analysis is a fairly recent analytical field that searches data for the likelihood of one variable actually causing another (as opposed to conventional statistics, which only look at correlation). *The Book of Why* (2018) by Judea Pearl has a good layman's description of how this works.

Several research questions were posed and answered, first using conventional numerical analysis, and second using causal analysis. While the results in general were clear-cut, some aspects are more difficult to explain.

Bottom Line

Four factors are shown to be causal to membership status, meaning 4 factors caused members to retain or drop membership. Those factors are *having a doctorate*, *being active on a committee or working group*, having an *INCOSE certification* (all of which are shown by both analysis of numbers of members who retain membership) and *gender* (which the analysis of numbers of members did not show affected retention).

Summary

Causal analysis is performed on membership data using characteristics of the data that is fed into three different algorithms, which use different methods to identify causality and are based on different assumptions. Depending on the amount of data and how strong is the causality, the methods can produce different results. Hence if different algorithms all show causality, it is highly likely that there actually is causality. For this reason four algorithms were run. They are called PC (for the inventor's names), FGES (for fast greedy equivalent search), and FCI (for fast causal inference). FGES has an option to run with different criteria for scoring graphs, based on assumptions. We used two, one called DBIC (digital binary information criterion) and one called BDEU (Bayesian Dirichlet-likelihood equivalence and uniform).

Two of these methods determined that all four factors above are causal to membership status. All four methods showed Certification status and Gender as being causal. Two methods showed uncertainty in whether activity in a committee was caused by, or caused, INCOSE membership status. One of those also showed uncertainty in whether INCOSE certification caused INCOSE membership status or vice versa, and one method showed causality in the opposite direction (Membership status caused certification).

In addition to the causal analysis, numerical analysis was also performed. 98% of members whose certification status was “in progress” retained membership, and 93% of members who have a certification retained membership, compared to much smaller numbers of members who are or were not certified.

74% of members who had any notation in the “Active Committees” column retained membership in recent years compared to 48% of members with blanks in that column. 62% of members with doctorates retained membership compared to 46% of members without doctorates.

Gender is an interesting result. While the causal analysis shows definite causality (gender causes membership retention or membership lapse), the fraction of male and female members since 2014 who have retained is essential equal, and not far apart for earlier years. It is possible that women felt less welcome in earlier years and were retained less, and more recently women were retained more; this would cause the numbers to be equal but still show causality. The actual reason is unanswered in this data.

Dataset Provided

Original member data included the following columns per member. (There were 55,541 members (rows) included.)

Member number, including 1 to 273709.

Salutation. This column included common salutations such as Dr., Mr., M., Mme., Ms., Prof., but also military titles such as Capt., Col., Major, variations such as Professor, non-English salutations such as Dipl.Ing., suffixes like Jr, III, or IV, letters signifying degrees such as C.Eng. or Ph.D., unidentified titles such as Barten, FBR, Herchert, or Sharma, special titles such as His Excellency, and even apparent phone numbers. 30,947 members had a blank in this column. The only way Salutation was involved in the analysis is if it indicated gender or whether the member had a doctorate (see below for examples).

Title. This included a wide variety of titles, from 2nd Year Student to Professor to Director to Retired, from Assoc. Professor in Systems Engineering to General Manager, from Academic to Six Sigma Black Belt to Adv Dev Lead to Software developer. Also included were addresses and unexplained numbers like 100081. There were 22,854 blanks. The only way Title was involved in the analysis is if it indicated gender or whether the member had a doctorate.

Member Type. There were ten member types as shown below. It seems that “Lifetime” refers to award recipients, who need not even have joined INCOSE to be included. Associate members are employees of INCOSE Corporate Advisory Board organizations who are not individual members of INCOSE. MOA means there is a memorandum of agreement between INCOSE and a systems engineering organization in a non-US country, and these are members of that organization. Senior members are at least 64 years old and have been individual INCOSE members for five years. Senior members, students and members from developing countries (designated by the INCOSE president*) are offered reduced membership rates. (Source: https://www.incose.org/about-incose/incose-membership/membership-types.)

#	Member Type
30,071	Regular
508	Senior Member
3343	Student
26	Lifetime
1122	Developing Country
6459	Associate Member
10,959	MOA Member
1515	MOA Developing Country
138	MOA Senior Member
1397	MOA Student Member

*the addresses of "Developing Country" members include Algeria, Argentina, Australia, Bahrain, Belarus, Botswana, Brazil, Canada, Colombia, Costa Rica, Chile, China, Ecuador, France, Germany, Ghana, Honduras, Hungary, India, Indonesia, Iran, Israel, Jamaica, Japan, Jordan, Kenya, Korea, Lesotho, Malaysia, Mexico, Morocco, Netherlands, Nigeria, Pakistan, Peru, Philippines, Poland, Portugal, Qatar, Russian Federation, Rwanda, Saudi Arabia, South Africa, Spain, Sudan, Taiwan, Thailand, Tunisia, Turkey, Ukraine, UAE, UK, and US. (It should be noticed that all of the ones from UK, US, Canada Netherlands, Germany, and Israel have lapsed in membership, and some are shown as a member of a different country than their mailing address.)

There were no blanks in the Member Type column.

Membership status. Members are either Active, i.e. paid up (17164), or Lapsed (38376). There were no blanks.

State. This column was filled for members from the US and some other countries, including most members from Australia and Canada, and some members from other countries. States were not used in the analysis.

Chapter. These included all INCOSE chapters (some emerging) plus "Member at Large Region 1", Region 2, or Region 3. There were 680 blanks, all of whom had Lapsed membership status. 6440 members listed a chapter of CAB; these were all Associate members and none had a Sector listed. Chapters were not used in the analysis.

Sector. This was shown as I Americas, II EMEA, and III Asia-Oceania. Blanks were either CAB members or 19 students whose chapter is "Sigma Theta Mu", all of whom are from the US, 17 from Michigan. Sectors were not used in the analysis.

Most Recent Certification, *Is Certified*, and *Most Recent Certification Status*. Most recent certification shows whether the person has an ASEP, CSEP, or ESEP. "Is Certified" is yes, no, or blank if never tried. Most recent Certification Status includes the following:

#	**Certification Status**		
145	Initiated		
8	Application Submitted		
12	Request for additional info		
90	Waiting for References		
37	Queued for Review		
13	In Review		
47	Passed Review		
63	Application Denied		
189	Ready for Exam		
7	Exam Failed		
3	On Hold		
1593	Application Expired		
2	Transitioning		
2807	Certified		
1177	Certification Lapsed	Subtotal	6193
49334	Blanks (assume = never initiated)		

Join Date, *Reinstatement Date*, *Expiration date*. Join dates range from 12/1/1991 (member #1) to 6/21/2018; the latest date is probably just before the dataset was sent over from INCOSE to the analysts. Only 9945 had a Reinstatement date; presumably these members let their membership lapse and then were reinstated, although some show reinstatement dates within a week of Join date so this is puzzling. One member has a join date in 2010, a reinstatement date in 2035, and an expiration date in 2016. Expiration dates range from 5/30/1992 through 2060. (Future dates beyond 2021 are all for lifetime members). Reinstatement and expiration dates ended up not being used in the analysis in any way. Dates were used to identify whether lapsed members were recently lapsed (lapse 2015 or more recent) or longer-ago-lapsed (before 2015). Aside from that, there were too many questions to use the dates for the first round of analysis.

Age and *Date of Birth*. Only 22 people had an age listed; ages listed ranged from 21 to 81; 20 had a date of birth as well. These were not used in the analysis.

Active Committee(s). Some of the notations in this column are committees or working groups, like Risk Management or Agile Systems and Systems Engineering. Others say "Awards" or Fellows; both of which are bestowed upon a member rather than indicating a specific activity. Some listed a sector or a chapter. Some members had an address and/or company in the Active Committees field. Many had multiple committees. Only 2552 members had anything in this field. Those who did were considered "Active members" in the analysis (separate from "paid up" members, although the word "active" is also used to mean "paid up" in the dataset and possibly in some early analyses). The members with nothing in the field were considered not clearly an active member, noting some might well be active even without noting the activity in the INCOSE database. All "Lifetime" members had

Awards in this field except two: one listed 15 committees (so Awards may have been an oversight) and the other had a blank field. This field is used in the analysis to see whether member activity causes retention.

Gender. 2248 members listed a gender (only Male and Female are in the database). Some who listed their gender joined in 1991, but most joined in about 2010 or later. Of the ones who had a gender, 1979 listed Male and 269 listed female. (Note that there were other ways to determine gender of some members--including “Mr.” or “Ms.” in either the Salutation or the Title column, for example--so the number of males and females used to determine research question answers is larger than 1979 and 269, respectively.)

Research Questions

The following research questions were posed.

Do men stay (maintain membership) more than women?

Do members with doctorates stay members more than members without?

Do people who are active (i.e. list anything in the Active Committee(s) field) stay members more than those who are not shown as active by that column?

Are members with doctorates more likely to be active? (This was posed because one of the possible causal diagrams suggests that “interest in subject,” which is not measured, may be a common cause of both having a doctorate and being actively on a committee, thus being a confounder.)

Do people who have INCOSE certification, or have certification in progress, stay members longer than those without?

The answers to these questions are shown in Table 1, and also in the “Calculations” tab of the Excel file.

Table 1. Answers to Research Questions

Question	Ans.	From	Notes
1. Does gender matter to retention?	Not according to this analysis	Comparing numbers	Numerical ratios of total numbers paid up vs lapsed are practically identical: Paid up is 47.4 +/- 0.3 % of total of (paid up + recent lapse), 10-11% of (paid up + all lapsed) for both genders. Note, however, that the causal analysis (line 6 of this table) did identify causality. This might suggest the numbers from different groups with different results chanced to cancel out. E.g. if earlier members who were female experienced left disproportionately, but later female members/student members stayed disproportionately more and canceled out the early losses, that might explain it.

Question	Ans.	From	Notes
2. Does gender matter to retention?	Yes	Causal analysis	All four causal algorithms showed that gender caused membership status (retention). The direction of the causality (i.e. male means more retention?) is not determined in a causal analysis, just that there is causality. Structural equation modeling (see future recommendations at the end of this report) may provide that direction, but this has not been done.
3. Are people with doctorates more likely to stay members?	Yes	Comparing numbers	The fraction of members with doctorates who reup (compared to recent lapses) is 33% higher than those not known to have doctoral degrees; almost twice as high when comparing those with doctorates who reup to all lapsed members.
4. Are people with doctorates more likely to stay members?	Yes	Causal analysis	All four causal algorithms showed that having a doctorate caused membership status (retention).
5. Do active members stay more than non-active members?	Yes	Comparing numbers	74% of those who have anything in the active committee column reup, compared to 48% of those who do not reup (using recent lapse comparison). Compared to all lapsed members, the ratios are 62% and 15%. Caveat: It might well be that lapsed members do not have anything in the Active column because they are lapsed, rather than the other way around. Caveat: It could also be that as members re-up time and time again, they start to be interested in being active, i.e. membership retention causing active interest as well as the other way around.
6. Do active members stay more than non-active members?	Yes	Causal analysis	Two algorithms showed causality in the direction of activity causing retention; the other two methods showed causality but were unable to determine direction
7. Are people with doctorates more likely to be active in INCOSE?	Yes	Comparing numbers	40% of those with doctorates have something in the Active Committee column, whereas only 24% who are not doctors and 15% of the ones whose doctoral status is not known have something in the Active Committee column.
8. Are people with doctorates more likely to be active in INCOSE?	Yes	Causal analysis	All four algorithms showed having a doctorate causes activity in INCOSE. The algorithm (FCI) that can determine whether or not there are confounders (such as interest in a subject causing both) showed that having a doctorate directly causes activity, it isn’t purely due to both factors having a common cause.
9. Does certification mean members stay longer?	Yes	Comparing numbers	This is the most obvious “YES” of this study. Paid-up members with INCOSE certification comprise 93% of (paid up + recent lapse) and 91% of (paid up + all lapse), whereas the numbers for those whose application expired, were denied or failed the test, or whose certification lapsed are less than 24%, and those who never sought certification are only 44% of (paid up + recent lapse) and 12% of (paid up + all lapse). Those who are partway through the certification process reup at an astonishing 98% rate!
10. Does certification mean members stay longer?	Yes	Causal analysis	Two algorithms showed causality (certification causing retention); one algorithm showed causality but was unable to determine direction, and the other algorithm showed causality in the opposite direction (INCOSE membership status causing certification status).

Question	Ans.	From	Notes
11. Are people with doctorates more likely to get certified? (This relationship was not studied numerically, but appeared when causal analysis was performed.)	Probably	Causal Analysis	Three algorithms showed having a doctorate causes getting certified (and the one that can tell if there are confounders showed there are no confounders), but one algorithm (FGES with DBIC) did not find a causal link.

Data Manipulation

Based on the research questions, code columns were set up. The figure below shows what conditions got what codes.

Codes		Stu	MbrCode	MemCode2	DrCode	Active Code	CertCode	GenCode
0		Not			Not	No		
1		MoA Student	RegSr PaidUp		Is Dr.	Yes	Cert	
2		Student Mbr	Recent Lapse				Partway Thru	
3		Reg Mbr Student	Lapse <2015				Appl Expired	
4		Sr Mbr Student	Not Reg/Sr				Denied /Failed	
5		AssocMbr Student		Assoc			Never	
6		DevCntry Stu		Devel Country			Lapsed	
7				MOA Devel Country				
8				Reg				
9				MOA				
10				Sr				
11				MOA Sr				
12				Student				
13				MOA Student				
14				Lifetime				
Blank		Unclear			Unclear	Lifetime Mbr, not active		Unclear
M								Male
F								Female
		Red Bold Text	(in data sheets)		**My Codes**			
		Green Bold Text	(in data sheets)		**My Codes, pasted as Values only**			

STU (student) was taken directly from the Member Code plus any other indication that the member is a student (e.g. put “Student” into title).

MbrCode (member code) had two parts. First, those who were not regular or senior members were given a 4, as it was felt that students, associate members, and those in developing or MOA countries might have different reasons for lapsing than regular members. Second, active members were given a 1 for “Paid up” and lapsed members were given a 2 (Recent lapse) if the lapse was 2015 or later) and a 3 if the lapse was before 2015). **Note that this means that all results below refer to regular or senior members and not to students, associate members, those from developing countries (regular or MOA), or lifetime members.**

MemCode2 (member code #2) separated out all the different kinds of members into different categories. As it turned out, this was not useful in causal analysis because the categories were not in any order. (A combined MbrCode that included 1-3 of Mbrcode plus 5-14 of MemCode2 was tried before separating out the two variables; this messed up causality as well, again because the categories were not ordered, hence the two variables were separated for the final analysis.)

DrCode (doctor code). Based on title, salutation, and occasionally, job (e.g. Professor), as well as whether or not the member is a student, some members were given a DrCode of 1 for having a doctorate or of 0 for not having a doctorate. Those whose status could not be determined were left blank.

ActiveCode (active-on-committees code). This was given a 1 if the Active Committee(s) column had anything in it and 0 if it was blank, except for lifetime members (award recipients). These were given a blank ActiveCode. As there were only 26 of these they were not considered important to study at this phase.

CertCode (certification status code) was based on the Most Recent Certification Status column. “Partway thru” (2) was given to all statuses other than certified (1), application expired (3), application denied (4), exam failed (4), certification lapsed (6), or never (5).

GenCode (gender code). Members who had an M or F in the Gender column original data set were given that designation in GenCode. The Title and Salutation columns were searched for indication of gender such as “Mr.” or “His excellency” or “Ms.” “M.” was considered masculine since the members using that as a title were generally from French-speaking countries. First names could not be used to identify gender, since names were redacted before we got them.

Comparing Numbers

As a preliminary look at the data, numbers of members who met certain criteria were counted. To do this, the analyst used the Excel function of Data Filtering (to select only members who had a certain value in a column of interest) and then Count (on the bottom bar).

Question 1: Effect of gender on retention

Q1 Do men stay more?		NO, same	GenCode	GenCode	GenCode	GenCode	
			M	F	M	F	blank
			Male	Female	Male	Female	Unclear
		n ->	19598	3425			32517
Mbrcode	1	Paid Up	1367	242			3848
Mbrcode	2	Recent Lapse	1491	271			3361
Mbrcode	3	Earlier Lapse			9056	1890	9055
Total			2858	513	11914	2403	
Fraction reupped			**0.478307**	**0.471735**	**0.1147**	**0.1007**	0.236596
		ratio		99%		88%	0.9863
				3.4		33.7	

Note that both genders had nearly identical fractions of members who are still paid up compared to the sum of "paid up" and "recent lapse" (47.8% for men, 47.2% for women).

Women's rate for staying paid up varies a bit more when all lapsed members are part of the denominator (men 11.5%, women 10.1%) This may have occurred if INCOSE was less welcoming to women at its beginning than it is now.

Of the members whose gender is unknown, 53.4% reupped (compared to recent lapses) or 23.7% reupped (compared to all lapsed members). These are higher than the numbers for members known to be either male or female. It is not clear why, though it may relate to incomplete record keeping early in INCOSE's history.

(The last line was a calculation to determine how close to exactly-the-same the numbers were. If only 3 more women (<1%) had stayed members in the last 4 years, out of 500, then the numbers would have been exactly equal. 33 (1.7% of 1890) would have had to stay members in the earlier years to be exactly equal. There is a hint here that female retention has improved from not-quite-equal to almost exactly equal.)

Question 2: Effect of doctorate on retention

Q2 Do Drs stay more?		YES	DrCode	DrCode	DrCode	DrCode	Unclear
			0	1	0	1	blank
			Not Dr	Dr	Not Dr	Dr	
		n->	36515	3039			15986
Mbrcode	1	Paid Up	2666	475			2316
Mbrcode	2	Recent Lapse	3088	297			1738
Mbrcode	3	Earlier Lapse			16662	1286	2053
Total			5754	772	22416	2058	
							0.571288
Fraction reupped			**0.46333**	**0.615285**	**0.118933**	**0.230807**	0.379237
		ratio		133%		194%	1.3280

3039, or approximately 1 in 13 of the members whose doctoral status could be determined, hold doctoral degrees. Of these, 475 are paid up compared to 297 that are recent lapses, meaning 61.5% of those with doctorates paid up (compared to recent lapses) (23.1% when compared to all lapsed members). 2666 of those who do not have doctorates were paid up, compared to 3088 who recently lapsed, for a 46.3% retention rate (or 11.9% compared to all lapsed members). Note those ratios are 33% and 94% higher for members with doctorates than members without.

For the 15986 members whose doctoral status is not known, the re-up rate is 57.1% compared to recent lapses and 37.9% compared to all lapsed members. The first number is between the numbers for those with doctoral degrees and those without, which is as expected. The number compared to all lapses is higher than either doctors or known non-doctors, which is a bit puzzling. This may have to do with less complete recordkeeping early in INCOSE's history.

Question 3: Effect of active committee work on retention

Q3. Do people who are actively on committees reup more?									YES BUT
					But: why are lapsed people listed as active anyway?				
					ActiveCode	ActiveCode	ActiveCode	ActiveCode	Lifetime
					0	1	0	1	
				n ->	52989	2529			22
	Mbrcode	1	Paid Up		4374	1083			
	Mbrcode	2	Recent Lapse		4748	375			
	Mbrcode	3	Earlier Lapse				19722	279	
	Total				9122	1458	28844	1737	
	Fraction reupped				**0.4795**	**0.742798**	**0.151643**	**0.623489**	
			ratio			155%		411%	

People without any active committees listed in the dataset column were only 48% likely to reup (compared to recent lapses only), or only 15.1% likely compared to all lapsed members, while people who did list active committees were 74.2% likely to reup compared to recently lapsed members and even 62% compared to all lapsed members (an astonishing quadrupling of retention for active members!)

There is a potential problem with this, though. Of all the members who lapsed long ago, why would any committees be listed as "active committees"? Are they actually active in the working groups as non-members, or not active and the indicator of activity was not removed? This data and conclusion may well be skewed by this issue.

Q4. Relationship of having a doctorate to being active on committees

This study was done to see whether there may be a hidden confounding variable, namely how interested the member is in the topic might be causing both having a doctorate and being activity in INCOSE.

Q4. Are people with doctorates more likely to be active?									YES
							Why are unclears so inactive		
					ActiveCode	ActiveCode			Fraction active
					0	1			100%
					Not Active	Active			
				n ->	52989	2529			
	DrCode	1	Dr	3039	287	188			40%
	DrCode	0	Not Dr	36515	2020	645			24%
	blank		Unclear	15986	9706	1695			15%

There are proportionally more members with doctorates who are active than members without doctorates who are active. (40% vs 24%). The caveat is that we would expect people whose doctoral status is unclear would fall between those two numbers, and there are proportionally less active than either. Perhaps this is related to the assigning of 0 as meaning not active; if some members simply have less information about them than others, then they would be considered in the Blank/Not Active cell and therefore skew the results.

Question 5. Effect of certification on retention

Q5. Do people who have certifications, or certs in progress, reup more?										YES
				MbrCo	MbrCode	MbrCode	MbrCode		Fraction reup	
				1	2	3	4		100%	
			n	Paid Up	Recent Lapse	Longer Lapse	Other		cf recent lapse	cf all
CertCode 1	Cert		2807	1704	125	35	943		93%	91%
CertCode 2	Partway*		544	331	7	1	205		98%	98%
CertCode 3	Appl. expired		2285	101	493	764	927		17%	7%
CertCode 4	Denied/failed**		70	8	26	16	20		24%	16%
CertCode 5	Never (blank)		49349	3246	4054	19185	22864		44%	12%
CertCode 6	Cert Lapsed		485	67	418	0	0		14%	14%

The highest retention of all (98%) is of people who have initiated but not completed the certification process. Second highest retention (93%) is people who have INCOSE certification. Those whose certification application derailed (the application expired or was denied, or they failed the test) and those who let their certification lapse have among the lowest retention (24% and 14%). (In the latter case, it's quite possible that they let their certification lapse because they decided not to be members anymore.) Those who have never initiated certification (44%) have less than half the retention of people who are certified or in the process of becoming certified.

Causal Analysis

Data that the causal analysis program (Tetrad, see http://www.phil.cmu.edu/tetrad/about.html) would be able to read was created from the provided data set. This meant keeping only the data to be analyzed (e.g., kept only MbrCode: 1=current,2=lapsed recently, or 3=lapsed before 2015; not actual date of lapsing). Also it involved removing everything from the Excel worksheet except the column headers and the values of the various codes, then saving the Excel file as a comma-separated-variables (CSV) file.

These Excel worksheets were created by copying the full data worksheet, saving all the values of the "Code" variables as values (rather than as calculations), then deleting all other columns.

This data was then put through the Tetrad program using several different algorithms. The algorithms included both major types (constraint-based—PC algorithm, Figure 1, and score-based—FGES algorithm, Figures 2 and 3, using different parameters DBIC vs BDeu, and FCI, Figure 4) to see to what extent the answers agreed. One algorithm (FCI algorithm, Figure 4) is able to determine in some cases that the causality is direct and is not the result of having unmeasured confounders.

The PC algorithm found all four factors (gender=GenCode, having a doctorate=DrCode, certification status=CertCode, and being active in INCOSE=ActiveCode) to have a causal relationship with membership status (MbrCode). It also found a causal relationship of undetermined direction between ActiveCode and CertCode, and it found that DrCode caused both CertCode and ActiveCode; and GenCode caused DrCode.

The two FGES runs produced similar but not identical results. Both showed causal relationships between all four factors and MemberCode, but the direction of causality was not determined in some cases, and in one case was opposite that shown by other methods, namely in Figure 2, MbrCode caused CertCode.

Both algorithms agreed with PC that GenCode causes DrCode, and DrCode causes ActiveCode. Also, GenCode causes CertCode. Aside from speculations below, there is not much information as to why these other relationships appeared.

FCI did not find clearly direct causality to MbrCode, but it did find that DrCode directly causes ActiveCode, and directly causes CertCode. In other words, FCI did find **direct** causal relationships (green bold arrows) between three of the variables that were examined as to whether they causally influenced member status. It also found that those three plus GenCode caused member status.

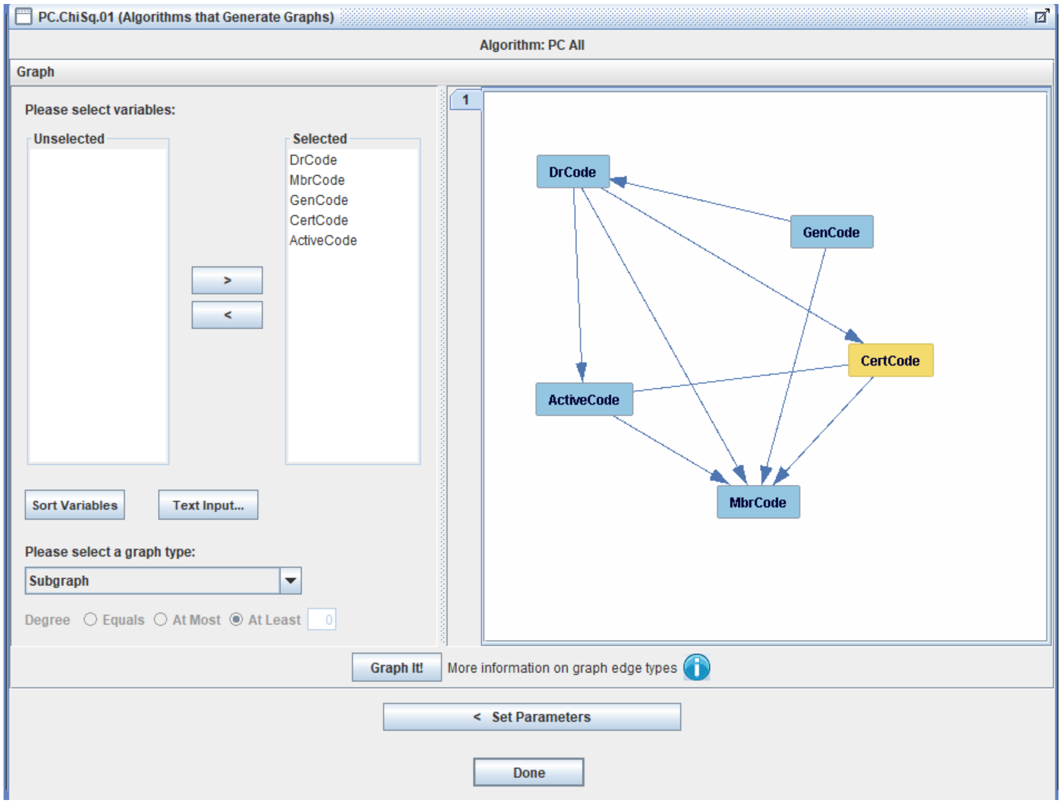

Figure 1. PC algorithm, using Chi Squared test and p-value 0.01

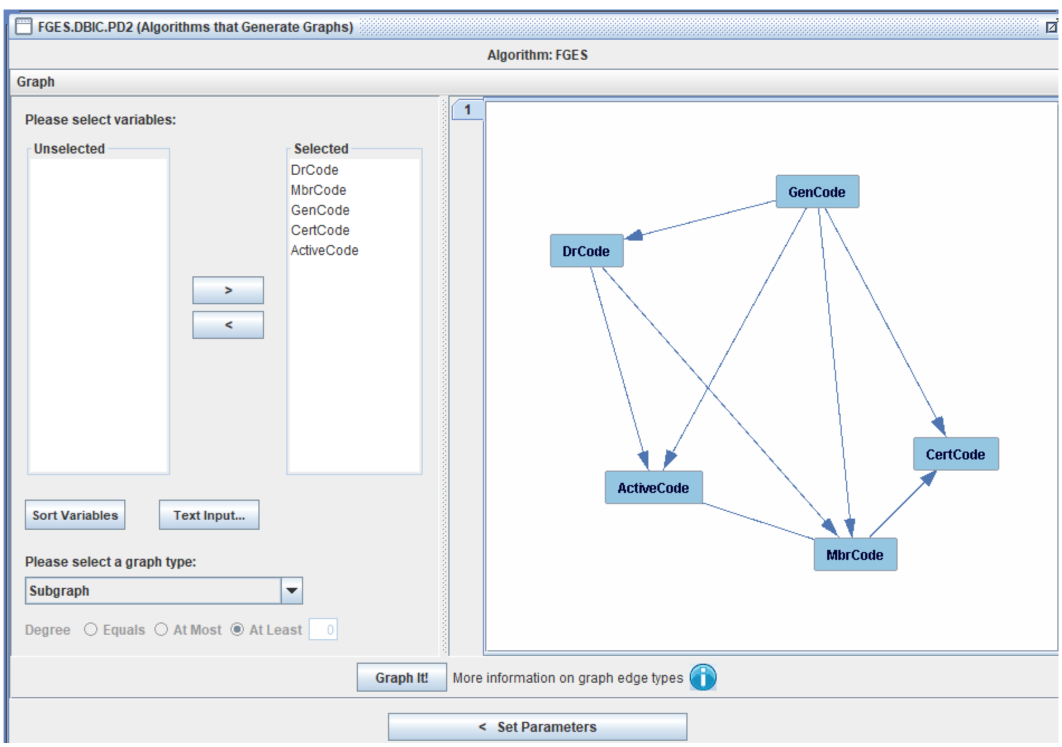

Figure 2. FGES algorithm, using dbic and pd 2

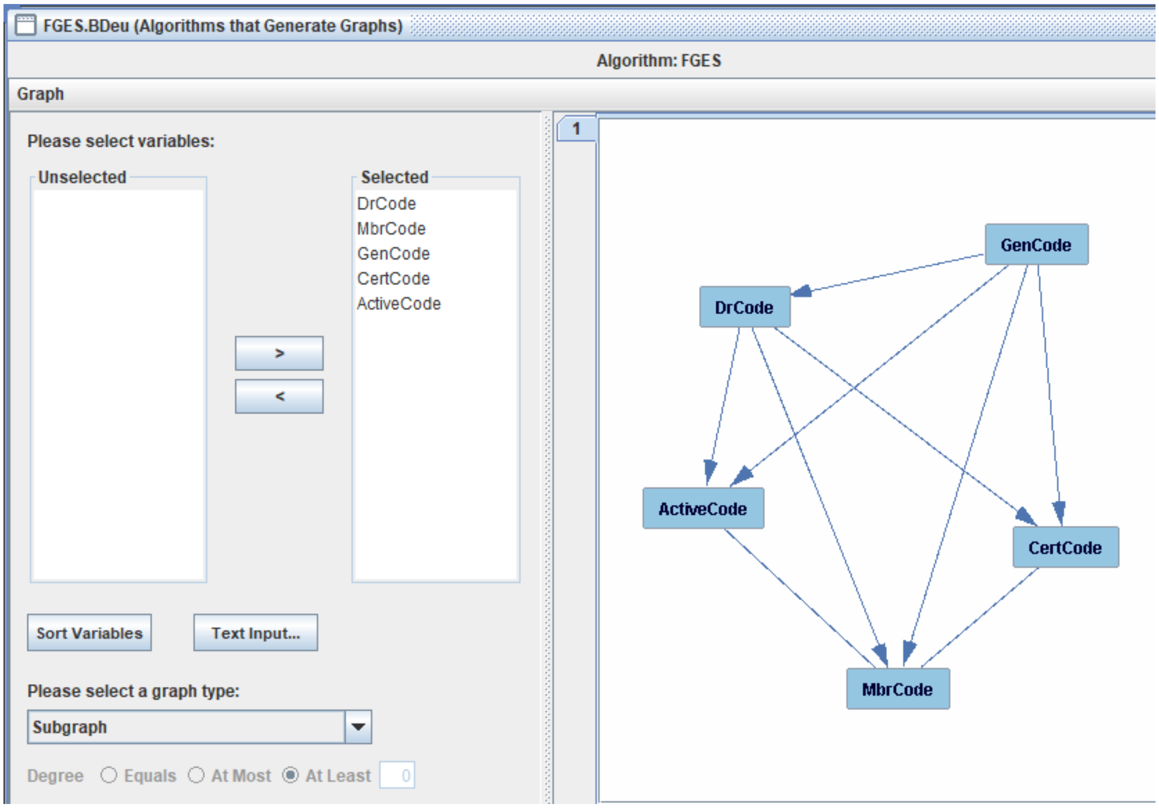

Figure 3. FGES algorithm using BDeu

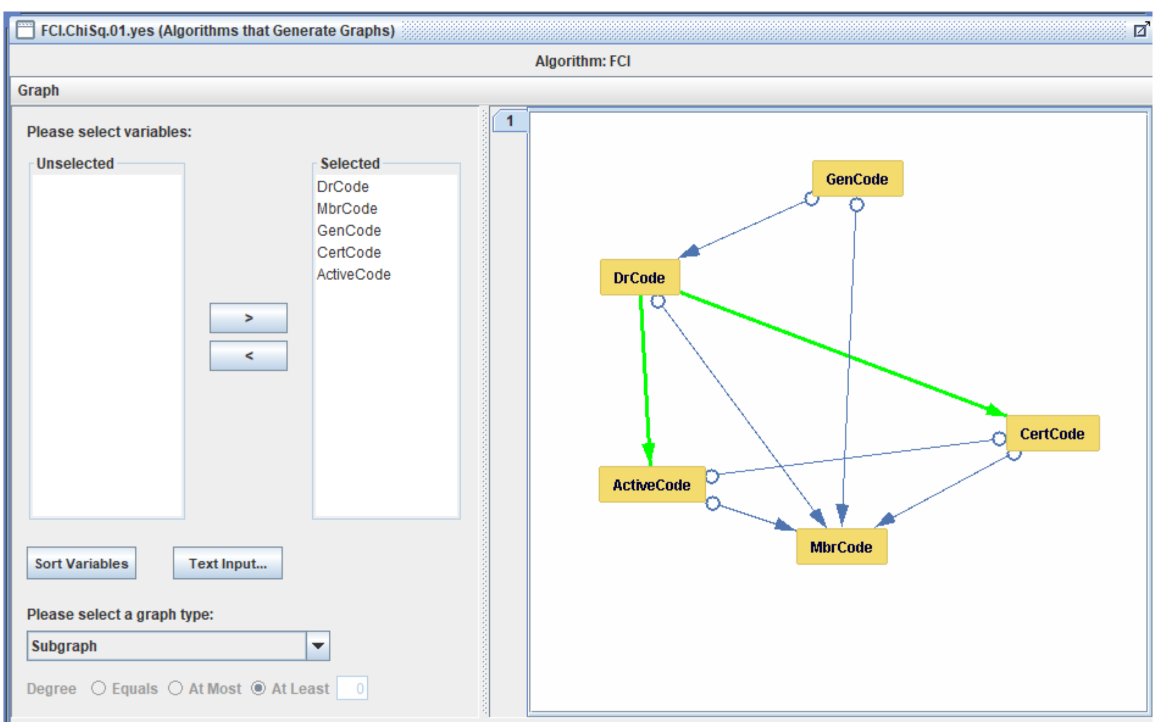

Figure 4. FCI algorithm, using ChiSquare, p=0.01

Causality Bottom Line

All 4 algorithms reported that DrCode and GenCode were causal on membership status. Meaning, whether you are male or female changes your likelihood of staying an INCOSE member, as does whether you have a doctorate or not.

Regarding ActiveCode, all four algorithms were showed causal relationships between ActiveCode and Member status, but two algorithms were unable to determine the direction. Of the two that did determine a direction, they found that *activity caused member status* and not the other way around.

Regarding Cert code, all four algorithms identified your certification status as having a causal relationship. One algorithm was unable to determine the direction of causality, and of the three algorithms that did find a direction, two found that *certification status caused membership status* rather than the other way around, and *one found the opposite*.

The most likely answer is that yes, all four factors do influence the likelihood that a member will renew and retain "active membership" status.

Three factors make sense given the numerical calculations in the previous section, further explained here:

- ActiveCode and DrCode. Assume some people are very interested in the topic of systems engineering or one of its subtopics. These people are likely to be active (ActiveCode) in an INCOSE working group or committee, and they may also be likely to have completed a doctorate in the subject (DrCode). Both being active and having a doctorate increase a member's chances of staying a member. (Note, however, that the FCI algorithm, which has the ability to distinguish between a direct causal effect and an apparent causal effect that is really caused by both factors having the same unmeasured cause, such as interest in the subject, showed that there *is* a direct relationship between these two. That does not rule out also having the same cause, though.)
- CertCode. Those who see having a certification as beneficial to their career are likely to be very motivated to continue to be involved with INCOSE. Moreover, those who have invested in the process already by filling out an application are the most likely to continue to be INCOSE members at least until they are officially certified (CertCode), at which point they become only slightly less likely to continue as members.

The fourth factor, gender, is more confusing. On the one hand, the numerical analysis shows no difference between proportions of males and females who re-up. However, all four causal analyses find that gender does influence the decision to stay a member. All four analyses also showed a causal relationship between gender and whether the member has a doctorate or not. The algorithms were split about whether gender influences activity in a working group or committee, or likelihood of having a certification.

The reason for the ambiguity about gender is not revealed by these methods. There might be a set of reasons that cancelled each other out in the overall numerical computations, yet showed up in the more complex causal analyses. For example, perhaps in the early days there were fewer women than men, and they were younger. Suppose also that they perceived sexism within INCOSE and left

INCOSE in greater percentages, but in recent years that problem has been reversed. Alternately, perhaps the age difference made women seem more likely to stay members (more men retired) but the fact that they perceived sexism made more drop out, and those two factors were not uncovered by the numerical totals. Part of the problem is that the *gender of few INCOSE members is known* for the earlier years. In this analysis, gender was not guessed from names (names were not provided), only from title/salutation when it was clear, and of course from the Gender column for more recent members.

Limitations

Membership status. Only regular or senior members were included in the analysis. It was felt that students would not be lapsing for the same reasons as regular members.

Algorithms. Not all possible search algorithms were tried, nor were all parameters that are input to the algorithms. As these make different assumptions, they can result in different answers. With the INCOSE dataset, the answers to the research questions were fairly clear, using algorithms both of the constraint type (PC) and of the score type (FGES and FCI), but it is possible that other algorithms could have come up with different answers.

Assumptions. Each algorithm makes assumptions about the data. We analyzed the data assuming factors were discrete rather than continuous variables (e.g., by putting the data into bins such as Currently a Member vs. Lapsed Recently vs Lapsed Longer Ago). Using a variety of algorithms helps offset the various kinds of assumptions (such as whether a true probability distribution is normal or skewed, and characteristics of noise), but more could be done to explore these.

Recommendations

First, it would be great if INCOSE could collect more consistent data on all members about the factors considered likely to cause member retention. This includes gender, possession of a doctoral degree, and more attention to activity, such as determining whether a member is active in a chapter, and re-surveying activity periodically, including by asking working groups and committee heads to state who is active in their groups.

Second, a “knowledge box” (restricting what factors could influence what other factors) was not used within Tetrad, but if one were used it might have helped the algorithms become more in agreement. For example, it is highly likely that the assignment of a person’s gender preceded any decision about INCOSE activity or the decision to obtain a doctorate. In causal analysis, such information about the situation can serve as guidance to both constraint-based and score-based algorithms.

Third, regarding the gender issue, INCOSE could consider setting up an anonymous survey of members that includes questions about perceived sexism or other possible reasons why one gender would be retained at a different rate than the other.

Fourth, studies could be done to see why students, associate members, or developing countries members (with or without MOAs) or regular or senior MOAs members lapse. The numbers in this analysis are only for regular or senior members who are not MOA members.

Fifth, studies could look at whether members in different countries, states, sectors, or chapters maintain membership at different rates and why. The Excel file (Calculations worksheet, rows 79-109) shows that countries definitely do maintain membership at different rates, varying from 0% (Brazil) to 97% (Australia). Determination of the reasons for these discrepancies would require additional data.

Sixth, some additional work could be done to understand the membership lapsing and reinstatement dates. These were confusing, as mentioned above, and thus they were not used in the analysis here.

Finally, a follow-up kind of analysis called structural equation modeling is available that can put numbers behind the arrows. So rather than saying the factor "has a causal relationship", you can say that having a doctorate makes you X percent more likely to retain membership.

Note that there is no funding to do any additional analysis.

Thank you for the opportunity to work with your datasets and provide you the answers about causality.

Contact Us

Software Engineering Institute
4500 Fifth Avenue, Pittsburgh, PA 15213-2612

Phone: 412/268.5800 | 888.201.4479
Web: www.sei.cmu.edu
Email: info@sei.cmu.edu

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM18-1374