Math Notes:
Pairing Vocabularies and the Design of Mathematical Sentences in Natural Languages

Jeff Cromwell, PhD

1 Abstract

Pairing of vocabularies of different organisms, scales and functions in a necessay function for mathematical sentences. Rules based on the pairing is the grammar in arrangment within sentences for the many to one relationship. Based on algorithm 1, several subgraphs were developed from the overall network G and mathematical sentences in English were generated based on subsets of the vocabulary in the form of three different data structures (a) Graph (b) Matrix and (c) English sentences for different numerical processing and categorification.

Keywords: Vocabulary Graph Theory Linear Algebra Coexpression Networks Oryza Sativa Auxin transporter-like protein 1

2 Introduction

The focus on the many to one relationship such as the many seeds contained in one seed, the attributes of a vocabulary with its pairing can provide insight when placed into one or collection of data structures such as the natural language sentence. Two vocabularies are developed at two different scalers such that (1) gene vocabulary from abstracts pairing of two organisms (2) molecular vocabulary from pairing two molecules based on molecular function based on many types and magnitudes of evidence. This conceptualization can be placed into other vocabularies like math as well with english words plus (i.e. +) and one (i.e. 1) in the pairing. Often, in applied work the english words for transcendental numbers such as e can be approximated with two point seven used in a sentence with the rules of grammar. For the number of combinations [1]

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \qquad (1)$$

For example, the pairwise combinations in Figure 1 N=930 for the vocabulary words and N=110 for the subset shown in Figure 3. Figure 1 has the coexpression network for LAX with graph G.
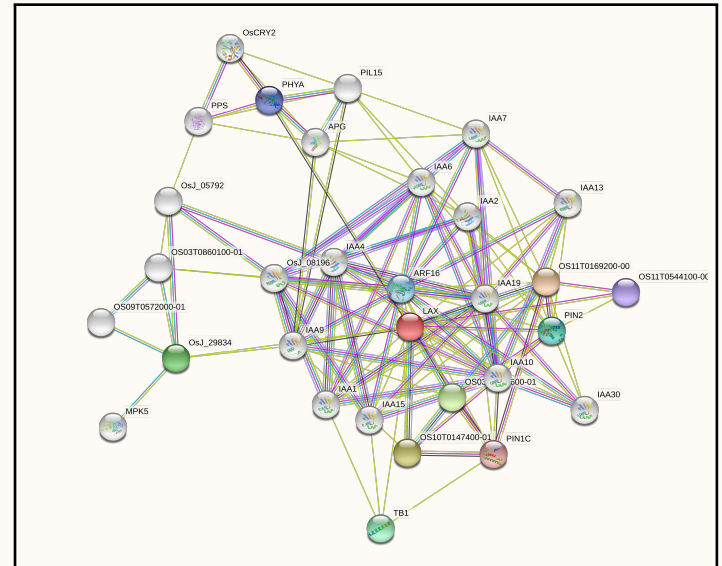

Figure 1: LAX Auxin transporter-like protein 1 Coexpression Network [601]

Table 1 has the vocabulary from the nodes of graph G.

1	APG	ARF16	IAA1	IAA10
2	IAA13	IAA15	IAA19	IAA2
3	IAA30	IAA4	IAA6	IAA7
4	IAA9	LAX	MPK5	OS03T0244600-01
5	OS03T0860100-01	OS09T0572000-01	OS10T0147400-01	OS11T0169200-00
6	OS11T0544100-00	OsCRY2	OsJ_05792	OsJ_08196
7	OsJ_29834	PHYA	PIL15	PIN1C
8	PIN2	PPS	TB1	

Table 1

Table 2 has the descriptions of each of the genes in Table 1.

	Molecule	Description
1	APG	Atypical bHLH transcription factor that acts as negative regulator of grain size. Binds the transcription factor ILI6 and forms a heterodimer of antagonistic bHLH transcription factors that regulates grain length and weight by controlling cell elongation in lemma and palea. May be involved in the control of lamina inclination through brassinosteroid signaling pathway
2	ARF16	Auxin response factors (ARFs) are transcriptional factors that bind specifically to the DNA sequence 5'-TGTCTC-3' found in the auxin-responsive promoter elements (AuxREs)
3	IAA1	Auxin-responsive protein IAA1; Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations
4	IAA10	Auxin-responsive protein IAA10; Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations; Belongs to the Aux/IAA family
5	IAA13	Auxin-responsive protein IAA13; Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations
6	IAA15	Auxin-responsive protein IAA15; Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations
7	IAA19	Auxin-responsive protein IAA19; Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations
8	IAA2	Auxin-responsive protein IAA2; Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations
9	IAA30	Auxin-responsive protein IAA30; Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations
10	IAA4	Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations
11	IAA6	Auxin-responsive protein IAA6; Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations
12	IAA7	Auxin-responsive protein IAA7; Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations
13	IAA9	Auxin-responsive protein IAA9; Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations
14	LAX	Auxin transporter-like protein 1; Transcription factor that seems to regulate organogenesis in postembryonic development. Involved in the regulation of shoot branching by controlling axillary meristem initiation. Functions in association with LAX2 to regulate the process of AM formation. Possesses transactivation activity in yeast; Belongs to the bHLH protein family
15	MPK5	Mitogen-activated protein kinase 5; Involved in disease resistance and abiotic stress tolerance signaling pathways. Acts as a positive regulator of drought, salt and cold tolerance. Negatively modulates pathogenesis-related (PR) gene expression and broad-spectrum disease resistance. Functions downstream of CPK18 in a signaling pathway that represses defense gene expression and negatively regulates resistance to rice blast fungus. Phosphorylated by CPK18 at Thr-14 and Thr-32 and activated independently of MAP kinase kinase (MKK) phosphorylation; Belongs to the protein kinase superfamily
16	OS03T0244600-01	AUX1 protein, putative, expressed; Os03g0244600 protein; cDNA clone:J033131K12, full insert sequence
17	OS03T0860100-01	cDNA clone:006-211-G07, full insert sequence; AP2 domain containing protein, expressed; Os03g0860100 protein; cDNA clone:006-211-G07, full insert sequence
18	OS09T0572000-01	Os09g0572000 protein; cDNA, clone: J065136G16, full insert sequence
19	OS10T0147400-01	Carrier protein involved in proton-driven auxin influx. May mediate the formation of auxin gradient from developing leaves (site of auxin biosynthesis) to tips (By similarity)
20	OS11T0169200-00	Carrier protein involved in proton-driven auxin influx. May mediate the formation of auxin gradient from developing leaves (site of auxin biosynthesis) to tips (By similarity)
21	OS11T0544100-00	Expressed protein; Os11g0544100 protein
22	OsCRY2	cDNA clone:J013042O16, full insert sequence; Putative cryptochrome 2
23	OsJ_05792	Os02g0200900 protein; Putative F-box protein
24	OsJ_08196	Auxin-responsive protein IAA8; Aux/IAA proteins are short-lived transcriptional factors that function as repressors of early auxin response genes at low auxin concentrations; Belongs to the Aux/IAA family
25	OsJ_29834	cDNA clone:J013047J08, full insert sequence
26	PHYA	Phytochrome A; Regulatory photoreceptor which exists in two forms that are reversibly interconvertible by light: the Pr form that absorbs maximally in the red region of the spectrum and the Pfr form that absorbs maximally in the far-red region. Photoconversion of Pr to Pfr induces an array of morphogenic responses, whereas reconversion of Pfr to Pr cancels the induction of those responses. Pfr controls the expression of a number of nuclear genes including those encoding the small subunit of ribulose- bisphosphate carboxylase, chlorophyll A/B binding protein, protochlorophyllide
27	PIL15	Os01g0286100 protein; cDNA clone:J033088H11, full insert sequence
28	PIN1C	Probable auxin efflux carrier component 1c; May act as a component of the auxin efflux carrier
29	PIN2	Auxin efflux carrier component 2; Acts as a component of the auxin efflux carrier. Involved in the basipetal polar auxin transport which contributes to the spreading growth of the tillers
30	PPS	cDNA clone:J013104O14, full insert sequence; Putative COP1, constitutive photomorphogenesis 1
31	TB1	Probable transcription factor that functions as a negative regulator of lateral branching, presumably through its expression in axillary buds . Involved in the fine tuning of shoot branching. May function as an integrator of multiple signaling pathways to regulate the development of axillary buds. Works partially downstream of strigolactones to inhibit bud outgrowth . Binds to MADS57 to suppress the negative regulation of D14 by MADS57 and balance the expression of D14 for tillering

Table 2: [601]

Table 2a and 2b have the amino acids counts for each of the sequences in Table 2.

	A	C	D	E	F	G	H	I	K
APG	89	6	36	15	14	28	20	5	9
ARF16	28	3	14	13	12	17	12	16	16
IAA1	32	1	19	10	8	8	1	1	12
IAA10	29	2	15	22	5	27	6	7	16
IAA13	39	1	14	17	6	21	0	4	25
IAA15	24	1	16	13	7	10	1	1	13
IAA19	38	6	20	21	4	25	2	3	17
IAA2	42	1	14	12	7	26	0	3	12
IAA30	46	3	27	17	6	28	1	7	25
IAA4	6	3	9	4	3	9	3	4	8
IAA6	28	4	22	23	10	28	5	12	25
IAA7	29	4	20	22	6	30	6	9	23
IAA9	17	4	19	9	5	21	4	2	11
LAX	49	9	11	16	37	41	15	31	18
MPK5	23	6	25	27	18	14	13	32	12
OS03T0244600-01	57	9	12	13	41	42	19	28	13
OS03T0860100-01	31	4	12	16	6	20	8	5	4
OS09T0572000-01	26	1	13	12	5	9	9	6	7
OS10T0147400-01	64	10	10	14	37	50	24	28	10
OS11T0169200-00	62	9	11	13	37	42	12	19	9
OS11T0544100-00	70	5	20	23	18	29	18	5	14
OsCRY2	40	12	41	49	21	37	19	31	30
OsJ_05792	73	33	32	34	15	54	6	24	29
OsJ_08196	17	6	11	8	4	18	7	5	8
OsJ_29834	42	4	49	41	12	38	18	36	40
PHYA	91	25	72	69	40	63	34	54	65
PIL15	93	9	34	31	29	72	17	19	21
PIN1C	67	4	24	21	30	43	9	38	20
PIN2	80	3	19	15	28	58	9	38	19
PPS	57	18	42	43	24	35	21	29	44
TB1	53	2	25	18	12	32	21	8	18

	L	M	N	P	Q	R	S	T	V	W	Y
APG	34	22	15	67	21	29	39	28	19	5	4
ARF16	37	6	8	22	14	12	34	14	16	4	11
IAA1	17	4	4	14	4	14	19	8	15	3	5
IAA10	21	7	10	14	12	17	28	19	16	4	4
IAA13	14	10	9	13	8	10	14	9	14	3	5
IAA15	18	5	5	13	3	14	35	10	15	3	5
IAA19	24	6	2	17	5	22	27	14	17	3	8
IAA2	17	7	5	17	5	15	31	4	12	3	5
IAA30	18	10	7	14	3	13	21	5	17	3	6
IAA4	14	6	0	6	4	12	5	4	9	2	4
IAA6	31	4	12	23	17	21	26	19	15	3	7
IAA7	27	6	11	14	7	17	30	16	14	4	5
IAA9	15	4	5	8	2	11	19	6	14	3	3
LAX	55	15	15	23	12	12	28	30	36	17	22
MPK5	36	17	21	24	13	27	13	16	15	2	15
OS03T0244600-01	55	11	16	32	12	23	30	36	38	16	21
OS03T0860100-01	15	11	4	11	11	20	24	7	12	2	2
OS09T0572000-01	20	10	4	18	13	19	41	9	9	2	5
OS10T0147400-01	59	11	14	33	17	21	36	32	38	16	23
OS11T0169200-00	43	13	14	21	9	27	24	31	47	16	21
OS11T0544100-00	55	10	2	67	13	26	51	15	36	8	4
OsCRY2	68	15	18	35	16	38	72	22	44	22	21
OsJ_05792	92	12	22	26	22	38	57	28	54	5	4
OsJ_08196	19	8	3	10	3	16	22	19	13	2	6
OsJ_29834	49	19	27	44	31	33	55	22	27	9	15
PHYA	121	36	45	45	48	54	90	48	92	11	25
PIL15	47	27	27	68	47	47	64	43	27	5	4
PIN1C	54	26	22	35	11	32	51	33	45	9	18
PIN2	59	24	23	42	17	30	63	27	49	9	18
PPS	55	17	29	29	28	38	66	32	50	7	21
TB1	21	12	12	30	18	23	44	12	16	3	8

Based on N Abstracts from Pubmed with keyword the Gene Symbol, Genes and Frequencies are provided in Table 2.

	Gene_symbol	Genes	Freq
1	T	T, brachyury homolog (mouse)	193
2	F2	coagulation factor II (thrombin)	139
3	C3	complement component 3	76
4	MYB	v-myb myeloblastosis viral oncogene homolog (avian)	70
5	GC	group-specific component (vitamin D binding protein)	55
6	TF	transferrin	51
7	CP	ceruloplasmin (ferroxidase)	40
8	CAT	catalase	37
9	RFT1	RFT1 homolog (S. cerevisiae)	33
10	F3	coagulation factor III (thromboplastin, tissue factor)	31
11	AR	androgen receptor	30
12	HR	hair growth associated	29
13	CS	citrate synthase	26
14	PC	pyruvate carboxylase	26
15	GCA	grancalcin, EF-hand calcium binding protein	24
16	SUB1	SUB1 homolog (S. cerevisiae)	19
17	HP	haptoglobin	18
18	PAM	peptidylglycine alpha-amidating monooxygenase	18
19	IMPACT	impact RWD domain protein	16
20	BRE	brain and reproductive organ-expressed (TNFRSF1A modulator)	14
21	C2	complement component 2	13
22	ERF	Ets2 repressor factor	13
23	ACE	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	11
24	GSR	glutathione reductase	11
25	MB	myoglobin	11

Table 3: [2], [802]

3 Algorithm

Algorithm 1 has the three different vocabularies based (a) organism A. (b) organism B and (c) math vocabularies with three different corpuses. [1000] Here then is a set of mathematical sentences in a language (i.e english) is developed from pairing the vocabulary. The mathematical sentences are composed from the sequence of graphics from the English Language.

Algorithm 1 A

1: Matchmaking with Keywords with a Corpus A
2: Generate a Vocabulary A in a Language A from Matchmaking
3: Matchmaking with Keywords with a Corpus B
4: Generate a Vocabulary B in a Language A from Matchmaking
5: Matchmaking with Keywords with a Corpus C
6: Generate a Vocabulary C in a Math Language from Matchmaking
7: For each Vocabulary A word there is a math sentence such that math symbol = is to the right of the vocabulary word.
8: For each Vocabulary Word A pair Vocabulary Word in A
9: If there is a relationship=1 else 0
10: If 1 then Add the Vocabulary word to the Math Sentence with number symbol multiplication symbol left of the word based on the Grammatrical Rules of the math sentence.
11: Solve the Equation System One with the Many Math Sentences in Language A

Code Listing 1 in the computer language R [1000] has an example of Algorithm 1. [800] [801]

```
1  algorithm.1<-function(x)
   {
   #API
4  require(combinat)
   require(stringr)
   require(stringi)
7  require(igraph)
   Network<-as.data.frame(read.delim(x))
   Network.Genes<-unique(union(Network$X.node1, Network$node2))
10 G1<-graph_from_data_frame(Network[,1:2])
   G1.ADJ<-as_adjacency_matrix(G1)
   xtable::xtable(matrix(Network.Genes,8,4,byrow = TRUE))
13 png(file = stringr::str_c(1,'.png'))
   op <- par(mfrow = c(1,1),mar=c(3,3,3,3))
   plot(G1)
16 par(op)
   dev.off()
   GS1<-subgraph(G1,Network.Genes[3:13])
19 png(file = stringr::str_c(2,'.png'))
   op <- par(mfrow = c(1,1),mar=c(3,3,3,3))
   plot(GS1)
22 par(op)
   dev.off()
   GS1.ADJ<-as_adjacency_matrix(GS1)
25 n<-nrow(GS1.ADJ)
   z1<-"";z2<-"";equations<-NULL;k=0;Z1A=""
   a<-1;b<-1;h=1
28 for(i in 1:n){
     z1<-stringr::str_c("d",rownames(GS1.ADJ)[i]," = ")
     for(j in 1:n){
31     if(GS1.ADJ[i,j]==1){k<-k+1
         z2<-stringr::str_c(z2, " + ",runif(1,a,b), "*",
             "f.",h,"(",colnames(GS1.ADJ)[j],")")
       }
34   }
     equations[i]<-stringr::str_c(z1,z2," + eps",i)
     z1<-"";z2<-"";k=0
37 }
   return(equation)
   }
```

4 Example

Consider the following example from Algorithm 1 where the 1 replaces One and + replaces Plus for the following mathematical sentences

where $f.1 = x^n/(1 + x)^n$, x is the vocabulary word and n is a positive integer exponent and eps is the environmental vocabulary word for possible treatment designs for each equation. Furthermore, transcendental numbers such as e can be approximated with two point seven used in a mathematical sentence with the rules of grammar.

1. dIAA1 = 1*f.1(IAA10) + 1*f.1(IAA19) + 1*f.1(IAA4) + 1*f.1(IAA6) + eps1
2. dIAA10 = 1*f.1(IAA1) + 1*f.1(IAA13) + 1*f.1(IAA15) + 1*f.1(IAA19) + 1*f.1(IAA2) + 1*f.1(IAA30) + 1*f.1(IAA4) + 1*f.1(IAA6) + 1*f.1(IAA7) + 1*f.1(IAA9) + eps2
3. dIAA13 = 1*f.1(IAA10) + 1*f.1(IAA19) + 1*f.1(IAA7) + eps3
4. dIAA15 = 1*f.1(IAA10) + 1*f.1(IAA19) + 1*f.1(IAA4) + eps4
5. dIAA19 = 1*f.1(IAA1) + 1*f.1(IAA10) + 1*f.1(IAA13) + 1*f.1(IAA15) + 1*f.1(IAA2) + 1*f.1(IAA30) + 1*f.1(IAA4) + 1*f.1(IAA6) + 1*f.1(IAA7) + eps5
6. dIAA2 = 1*f.1(IAA10) + 1*f.1(IAA19) + 1*f.1(IAA4) + eps6
7. dIAA30 = 1*f.1(IAA10) + 1*f.1(IAA19) + eps7
8. dIAA4 = 1*f.1(IAA1) + 1*f.1(IAA10) + 1*f.1(IAA15) + 1*f.1(IAA19) + 1*f.1(IAA2) + 1*f.1(IAA6) + 1*f.1(IAA7) + 1*f.1(IAA9) + eps8
9. dIAA6 = 1*f.1(IAA1) + 1*f.1(IAA10) + 1*f.1(IAA19) + 1*f.1(IAA4) + eps9
10. dIAA7 = 1*f.1(IAA10) + 1*f.1(IAA13) + 1*f.1(IAA19) + 1*f.1(IAA4) + eps10
11. dIAA9 = 1*f.1(IAA10) + 1*f.1(IAA4) + eps11

5 Results

Figure 1 has the graph G1 based on the pairwise relationships developed. [1005]

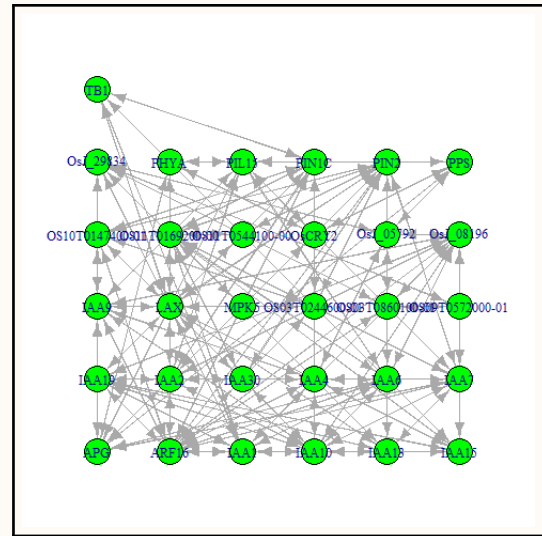Figure 2: Auxin transporter-like protein 1 [601]

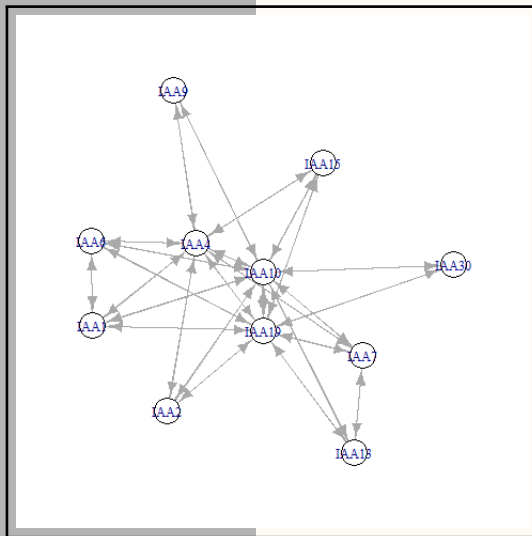Figure 2 has the subgraph G2 based on the Network subset j=3,4,...,13 from graph G1.[1005]

Figure 3: Auxin-responsive protein [1000][601]

6 Conclusions

Based on algorithm 1, a example subgraph was developed from the overall network G and mathematical sentences in the natural language English were generated based on subsets of the gene vocabulary from coexpression networks in the form of three different data structures (a) Graph (b) Matrix and (c) English sentences for different numerical processing algorithms and categorification. Here the letter and number in the symbol category are the same with respect to sequence formation with graphics or glyphs. The natural language composition of these graphics with engineering techniques with Oryza Sativa and Auxin transporter-like proteins can be helpful in the exploration of combinatorics in mathematical applications.

7 Bibliography

[1] Wikipedia contributors. "Combination." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 9 Nov. 2021. Web. 11 Nov. 2021.

[2] Wikipedia contributors. "PubMed." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 6 Oct. 2021. Web. 9 Nov. 2021.

[601] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-D613. doi:10.1093/nar/gky1131

[800] Cromwell, JB. (2021). The Mathematical Learning Space Research Portfolio:Main and Reader Pages. mathlearningspace.weebly.com. October 27,2021.

[801] Cromwell, JB. (2021). MathematicalLearningSpace. https://github.com/MathematicalLearningSpace October 27, 2021.

[802] Cromwell, JB (2021). A Review of Vocabularies for a Subset of the Gastric Cancer Literature for Years 2020-2021. Retrieved November 12 2021 at https://github.com/MathematicalLearningSpace

[1000] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[1003] Hans Werner Borchers (2017). pracma: Practical Numerical Math Functions. R package version 2.0.7. https://CRAN.R-project.org/package=pracma

[1005] Csardi G, Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. http://igraph.org