

LPI Alumni Skill Set Analysis



Research Question and Methodology

- What's special about LPI education? Did it manage to cultivate interdisciplinarity in graduates and how?
- Methodology:
 - Scrape LPI graduates LinkedIn profiles
 - Credits to jean-marc.sevin@learningplanetinstitute.org
 - Data cleaning and preprocessing
 - Get every profile's skillset and form network by co-occurrence
 - Cluster with networks
 - Analyze what's the key skill and how skills connect
 - Cluster by word embeddings
 - Analyze the differences



Raw Data

Name	Position	Company	Location	Experience	Education	Skills
a9240fad37c5b24f86528cb7a777af7d	MRes Biomedical engineering Research trainee...	CIML	Paris, Île-de-France, France	[['Stage de recherche en immunologie', 'CIML ...	[['Université Paris Cité', 'Master BME, Scienc...	[Organisation, Communication scientifique, Mic...
601d4f15cc5079606787744693d22bce	Business Operation Project Manager chez Hublo	Hublo	France	[['Business Operation Project Manager', 'Hublo...	[['Sorbonne Université', "Master Management de...	[Microsoft Office, slack, Organisé, Science, C...
b74c1a7cb052cccbc104ec2afe541aa4	M2 AIRE master (Life Sciences), iGEM Paris Bet...	Université de Paris	Paris, Île-de-France, France	[['M2 internship in Quantitative Evolutionary ...	[['Learning Plane Institute', 'Master AIRE Li...	[Anglais, Compétences en laboratoire, Anglais ...
a42d99b2480095a7cc4de90432dc47ed	MSc in Fintech, Finance & Digital Innovation -...	NaN	Milan, Lombardy, Italy	[['Country Business Development', 'Tovaritch I...	[['MIP Politecnico di Milano', "Master's degree...	[Travail d'équipe, Parler en public, Polyglott...
fd9e1c29af79ecdaedb5432dccb96368	Animateur en centre Fontenay-aux-roses	AgroParisTech - Institut des sciences et indus...	Antony, Île-de-France, France	[['Stagiaire', 'GreenBack', 'May 2021 - Jul 20...	[['AgroParisTech', 'Master 2 (M2), Sciences du...	[Music Theory, Piano Playing, Classical Ballet...

Problems:

- Both French and English
- Unstandardized user input

Data Pre-processing

- Data translation using Helsinki-NLP/opus-mt-fr-en pretrained translation model
- Skill name standardization by mapping into emsi skill name
 - EMSI skillset: An open-source library of 32,000+ skills gathered from hundreds of millions of online job postings, profiles, and resumes—updated every two weeks. <https://lightcast.io/open-skills>
 - Mapping relation is established by finding the closest l2 distance of sentence embeddings
 - sentence-transformers/all-MiniLM-L6-v2, SOTA sentence embedding model on huggingface
- Validation of cleaning is done by manual examination

Example of EMSI skill hierarchy

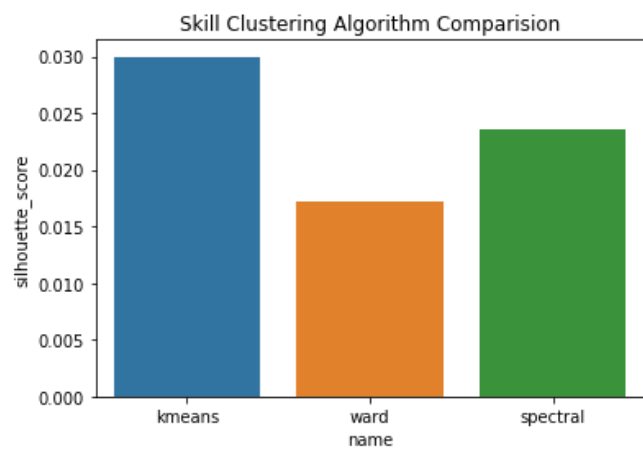
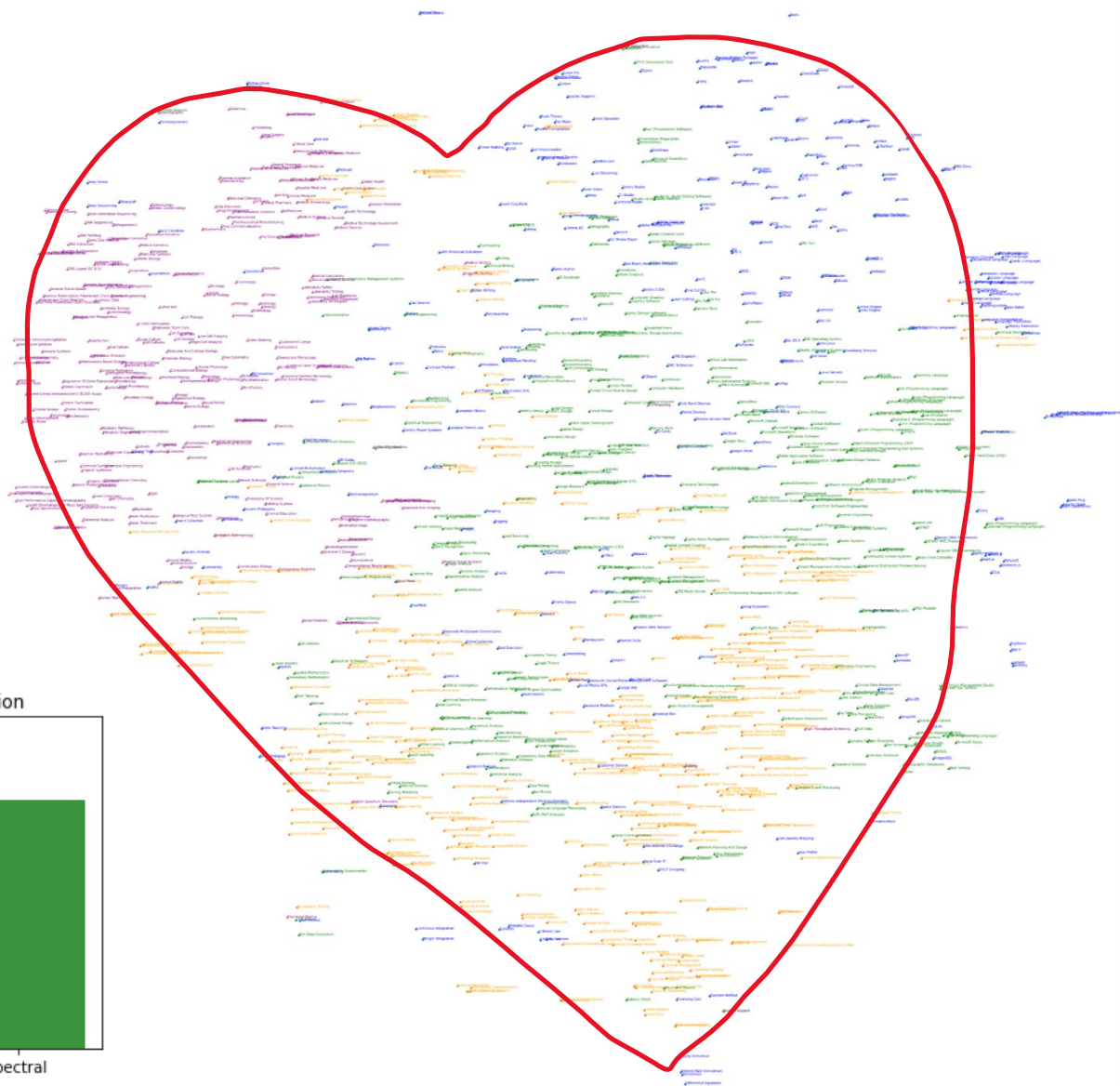
isLanguage	isSoftware	name	categoryName	subcategoryName	isHardSkill	isSoftSkill	wikipediaDefinition
False	True	.NET Assemblies	Information Technology	Microsoft Development Tools	True	False	\nDefined by Microsoft for use in recent versi...
False	False	.NET Development	Information Technology	Microsoft Development Tools	True	False	The .NET Framework is a software framework dev...
False	True	.NET Framework	Information Technology	Microsoft Development Tools	True	False	The .NET Framework is a software framework dev...

Example of Mapping Result

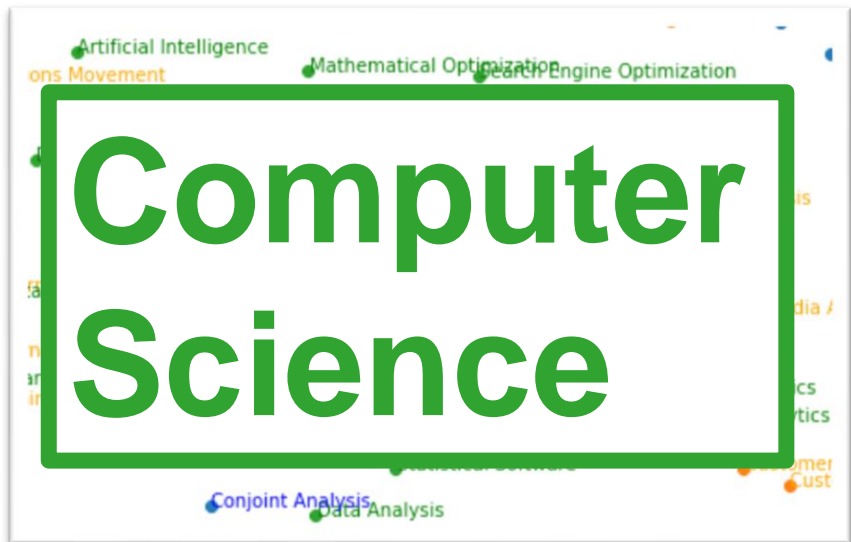
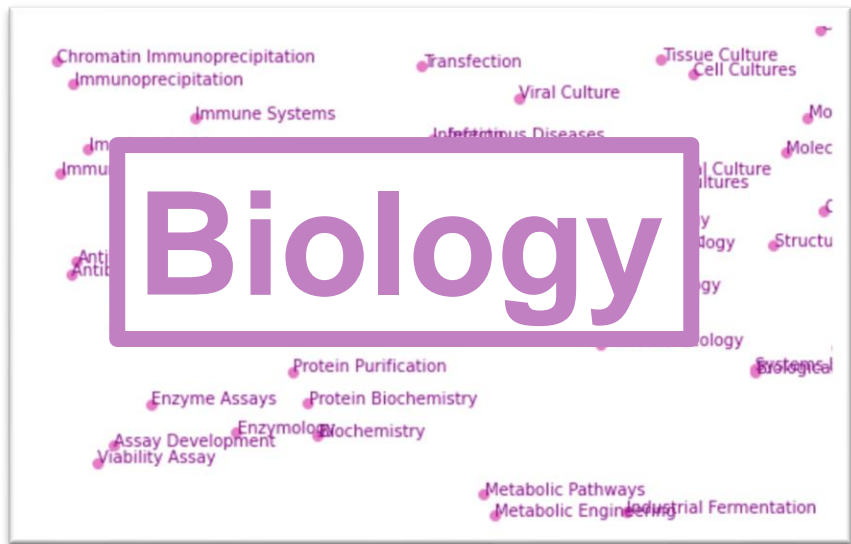
original_name	Literature Reviews	Computational Modeling	Animal Behavior	Python (Programming Language)	Research	Science	Organization Skills	Cognitive Neuroscience	Programming	Deep Learning
cleaned_name	Literature Reviews	Computational Modeling	Animal Science	Python (Programming Language)	Research	Physical Science	Administrative Skills	Cognitive Neuroscience	Computer Programming	Deep Learning
subcategoryName	Literature and Literary Studies	Mathematics and Mathematical Modeling	Biology	Scripting Languages	Research Methodology	Earth and Space Science	Administrative Support	Neurology	Computer Science	Artificial Intelligence and Machine Learning (...)
subcategoryName	Media and Writing	Analysis	Science and Research	Information Technology	Science and Research	Science and Research	Administration	Health Care	Information Technology	Information Technology

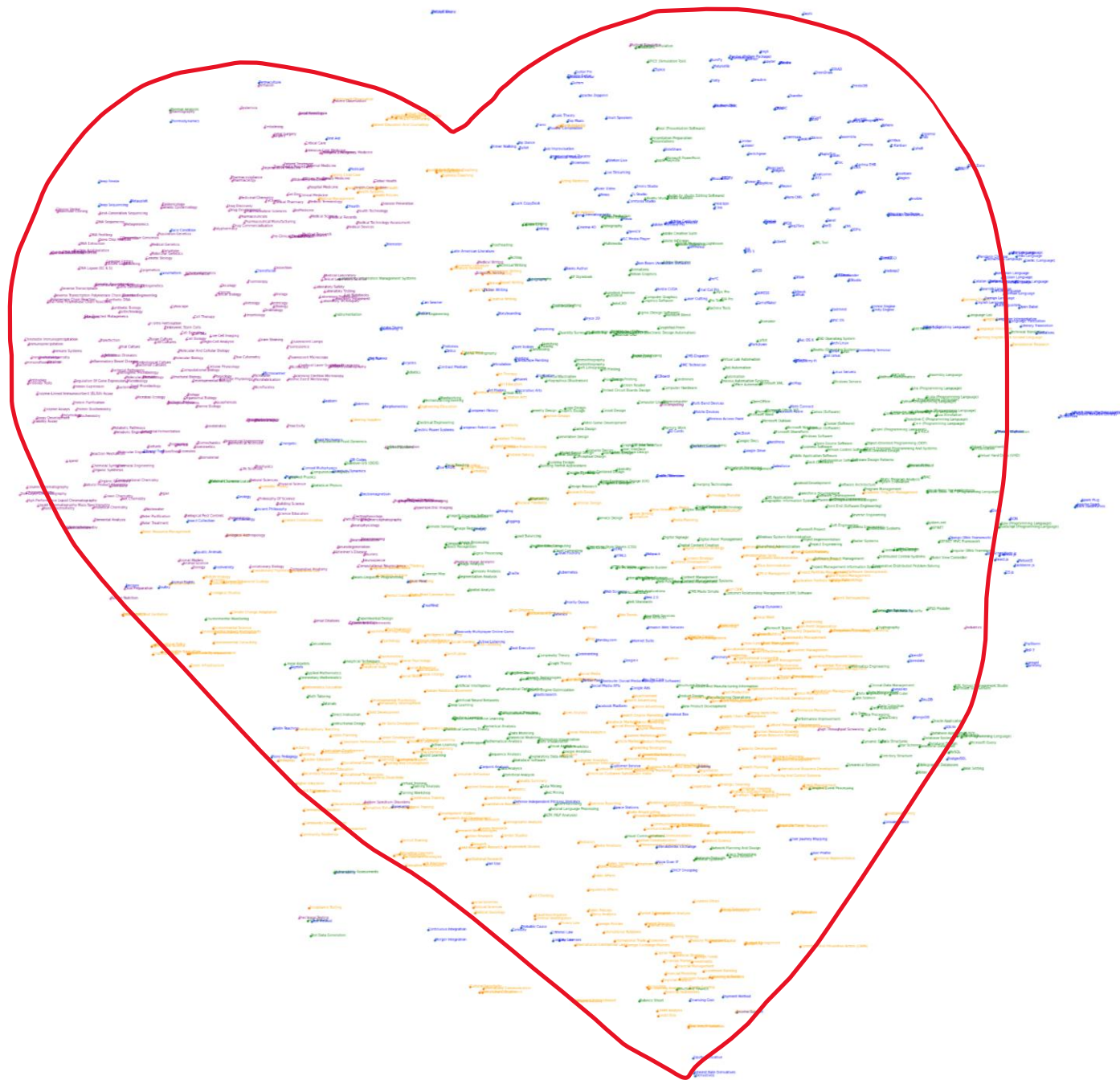
Results and Interpretations

Cluster of Skill Embedding



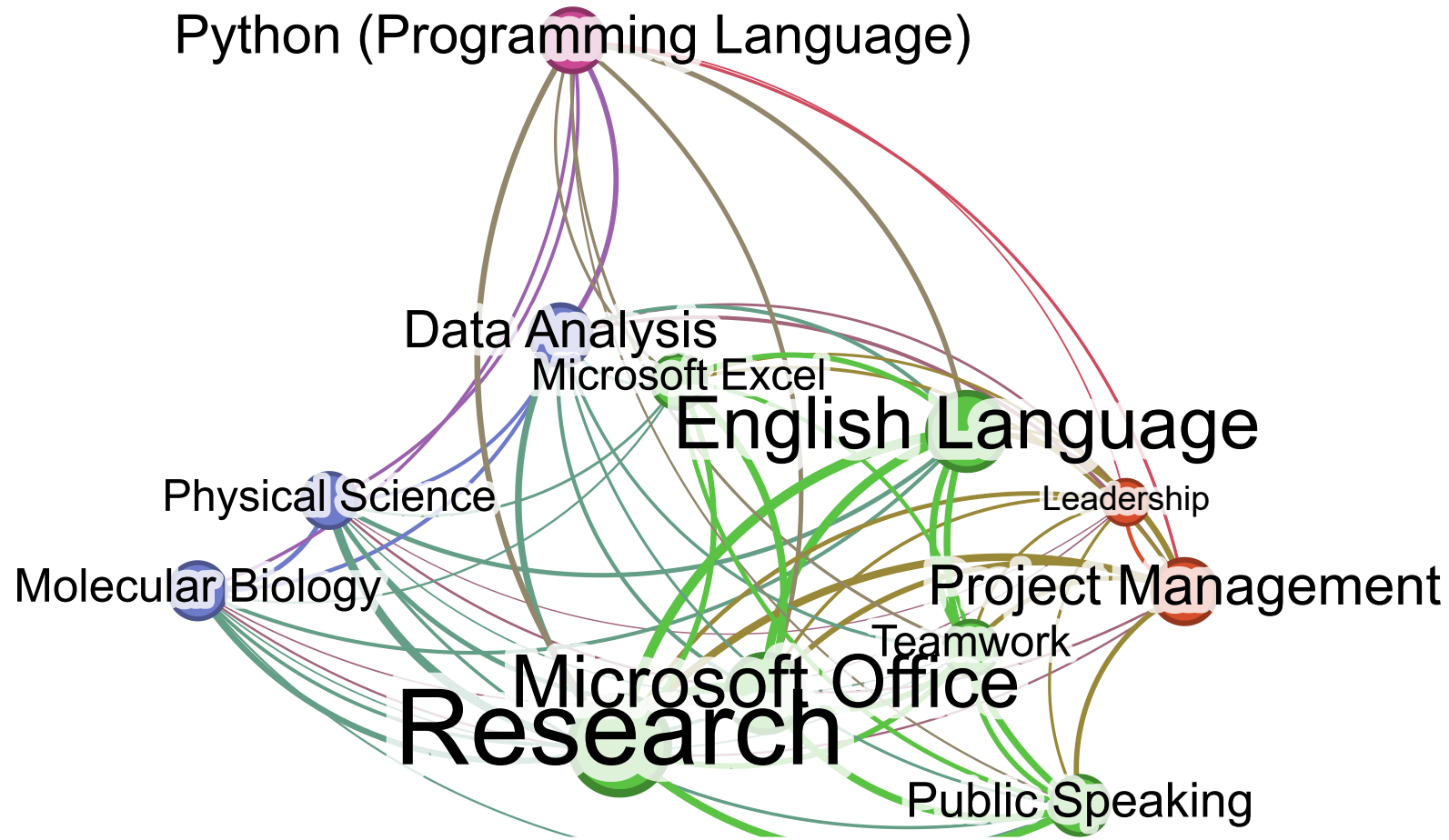
In details





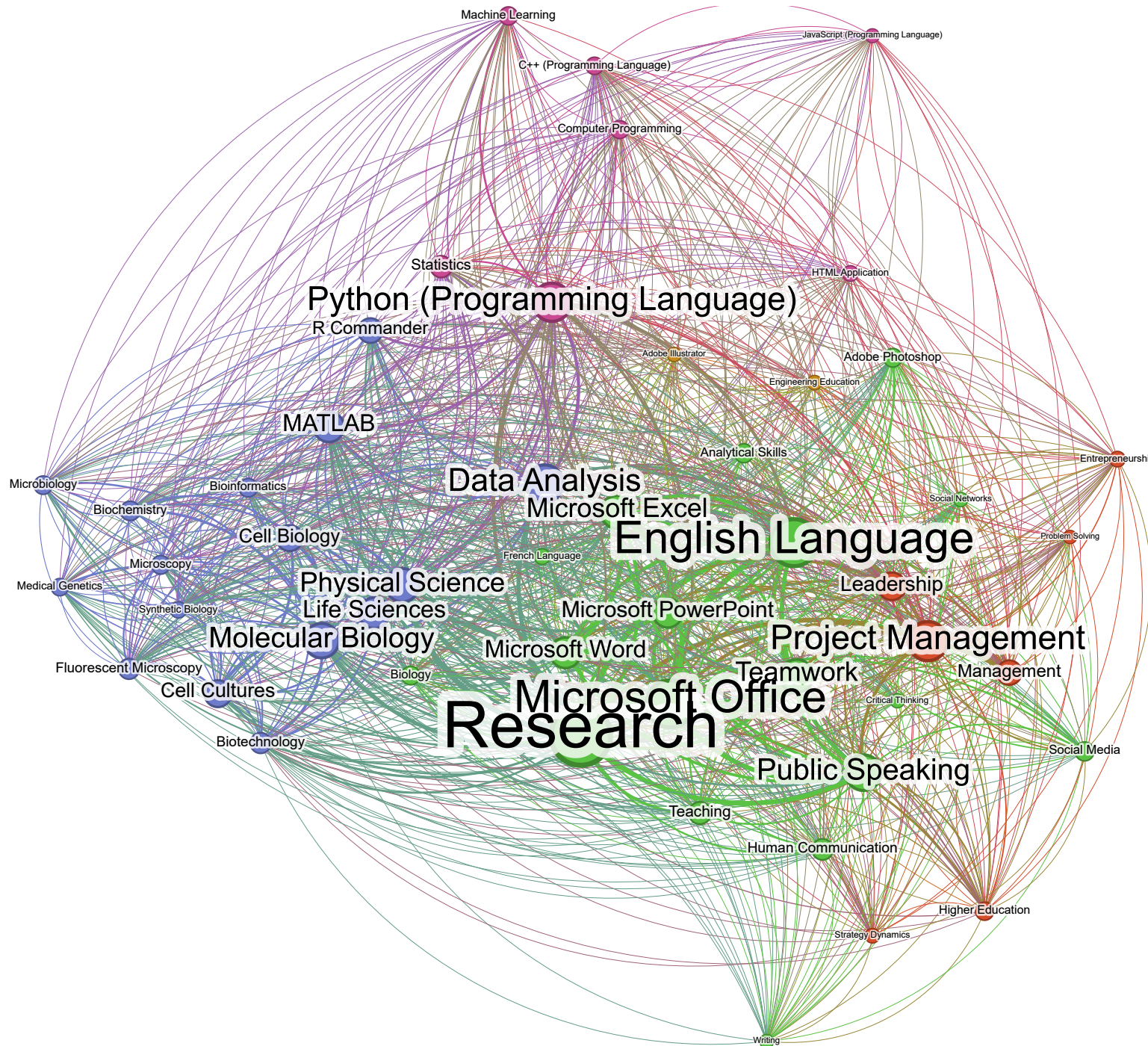
- In word embedding,
- Computer science and social science are more intertwined.
 - Biology is well separated

Skill Network based on co-occurrence



- Similar clusters also are present in network clusters.
- Soft skills being the core connecting all domains together.

In details



Conclusions:

- Soft skills being the core of interdisciplinarity
- Research is the main focus of LPI
- Biology is the most popular domain
- Ability to use Microsoft is stressed
- Python is the most popular programming language
- Learning science is disappearing or merging in?

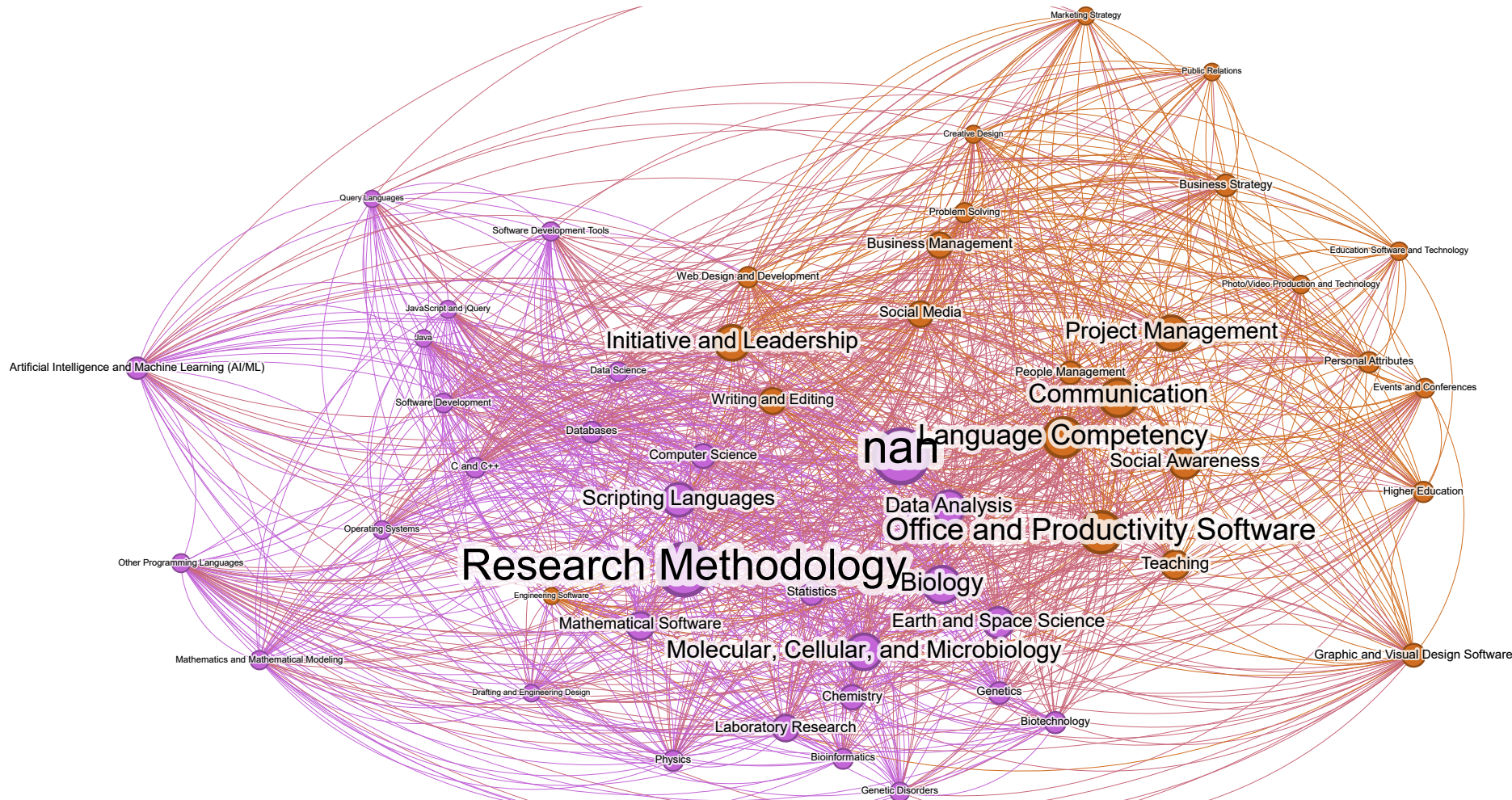
Limitations

- When analyzing the data, more background knowledge about LPI education is needed.
- Only part of the scrapped LinkedIn data has skill column. The data we use contain some licence and master profiles. Thus, the data is partial and our results may not be representative of LPI.
- When analyzing the network, due to computation resource and time limitation, we adopt gephi and use atlas 2 to do the clustering. It would be easier to analyze the results if more knowledge about the network clustering progress.

Thank you for listening 😊

Appendix

Subcategory Network based on co-occurrence



Category Network based on co-occurrence

