TheGraph.com Analysis Conducted by: 2/25/2023
Kevin Chu, Holly Cunefare, Todd Gardiner, & Yamineesh Kanaparthy

Research Question:

- Are the R&D Roadmap from TheGraph, year-end thesis from Messari, and the annual letter from Pantera Capital in strategic alignment?
- Can strategic alignment be proven with basic NLP methods applied to these corpora, and validated with publicly available chain data from TheGraph.com?

Original Documents:

- https://messari.io/pdf/messari-report-crypto-theses-for-2022.pdf
- https://panteracapital.com/blockchain-letter/the-year-ahead/
- https://thegraph.com/blog/roadmap-2022/

Python Output Word Cloud - Messari Thesis
Massarithesis.ipynb & massari.txt TheGraphroadmap.ipynb

Python Output Word Cloud - Pantera Capital Annual Letter
PanteraCapital.ipynb & pantera.txt

Python Output Word Cloud - TheGraph Roadmap, Z-Scores, Means, & Final Outputs
TheGraphroadmap.ipynb & graphroadmap.txt

Initial Result: The initial three word maps look strikingly similar (following page).

Further Steps: To determine the statistical significance we:

- Mean, and Z-scored all Word Map outputs to facilitate comparison
- Sum and Average Z-scores across the Word Maps
- Find the first statistically significant item in terms of joint popularity (Z-Scores > 4)
  The Top Scoring Word is **protocol** (Z-Scores > 5.6)
- Use TheGraph api to find corroborating information

[252 rows x 5 columns]

word	Pantera	Graph R&D	Messari	Mean
s	12.90	1.31	52.54	22.25
will	14.93	6.32	18.80	13.35
protocol	5.68	8.26	6.55	6.83
market	9.52	0.75	10.18	6.82
network	3.20	10.77	6.21	6.72
...	...	...	...	...
harder	-0.42	-0.36	-0.19	-0.32
touch	-0.42	-0.36	-0.19	-0.32
familiar	-0.42	-0.36	-0.19	-0.32
perform	-0.42	-0.36	-0.26	-0.34
automatical	-0.42	-0.36	-0.26	-0.34

(Messari Thesis Word Map)

(TheGraph R&D Word Map)

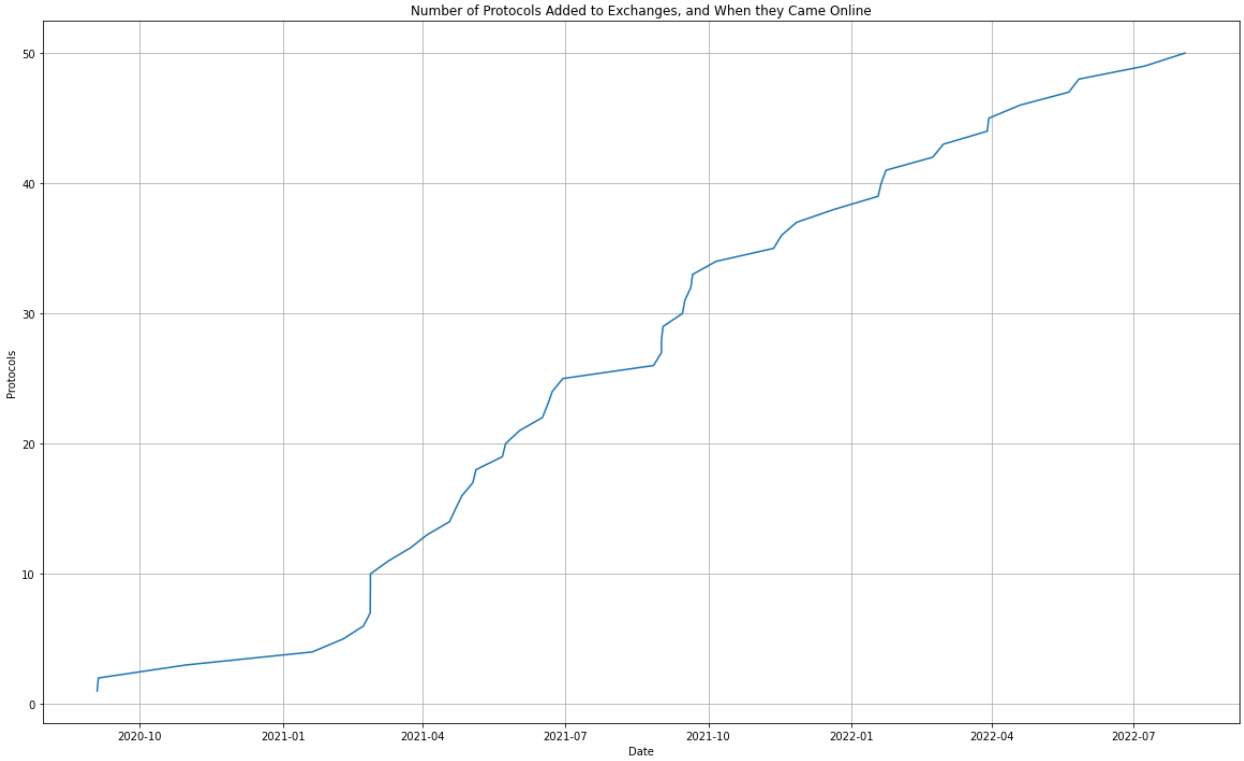
(Pantera Capital Word Map)

Do TheGraph Protocols added to the Marketplace support this conclusion?

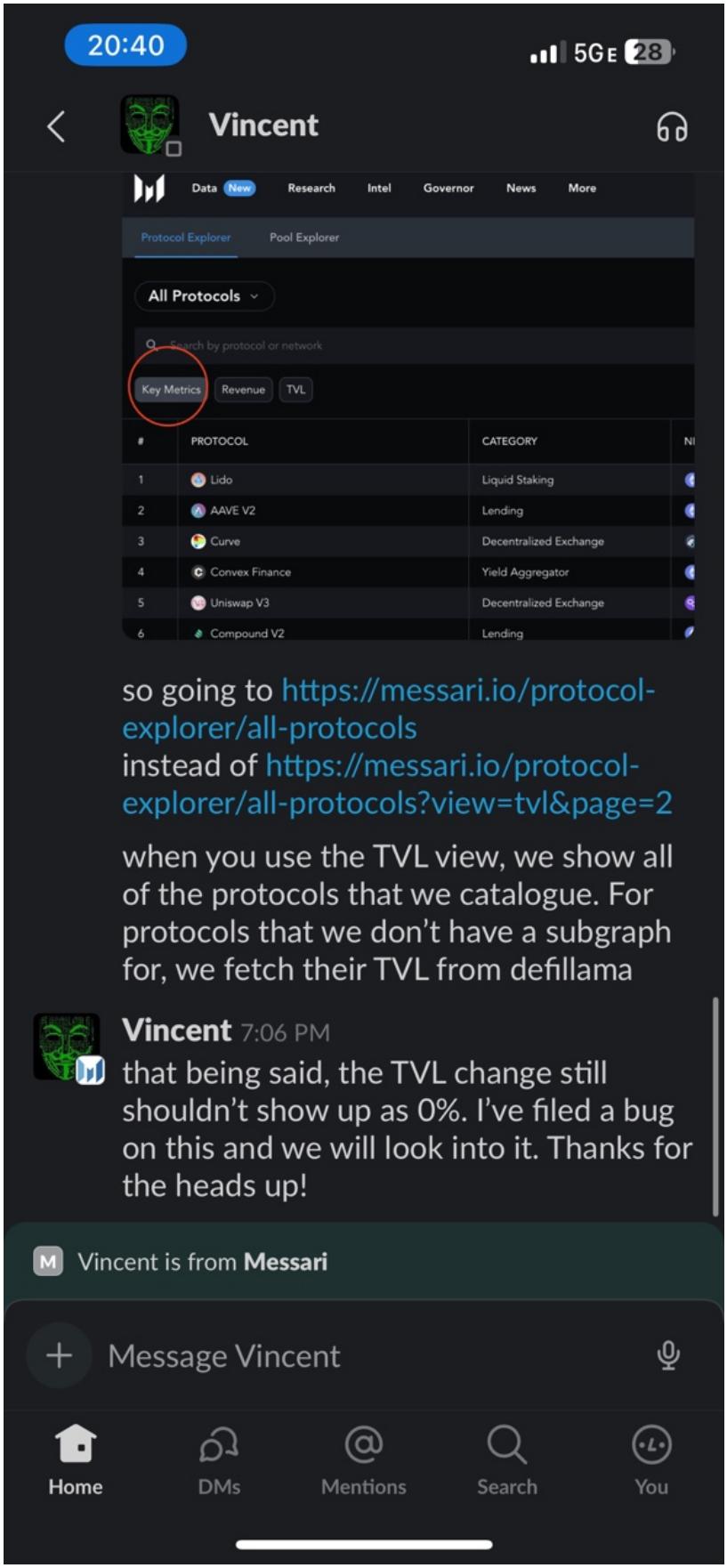
- When did the protocols come online?
  Yes, there has been a steady increase of protocols added to the marketplace section of TheGraph since September 2020. (See graph on the following page.)

- What is the value of all protocols that have come online? (https://subgraphs.messari.io/)
  Value of these protocols across the Exchanges Section of Messari Subgraphs: $6,892,935,189,886,955 (this may be a second bug to report)
  (Attempts to error check resulted in bug reported below)

Graphical Representation of Protocols added to Marketplace Since Sept. 2020



This plot shows a 100% increase in protocols added to the Exchange section of Messari Subgraphs in the first 4 months of 2021 (https://subgraphs.messari.io/). Subsequent doubling, 100% growth, is realized by July 1, 2021. The systems report further 50% growth in the latter half of 2021. These significant steps were followed by 40% YoY growth across the Messari Exchange Subgraphs in 2022, despite depreciation of multiple Version 2 systems.


20:40
5GE 28
Vincent
Data New
Research
Intel
Governor
News
More
Protocol Explorer
Pool Explorer
All Protocols
Search by protocol or network
Key Metrics
Revenue
TVL
PROTOCOL CATEGORY
1 Lido Liquid Staking
2 AAVE V2 Lending
3 Curve Decentralized Exchange
4 Convex Finance Yield Aggregator
5 Uniswap V3 Decentralized Exchange
6 Compound V2 Lending
so going to https://messari.io/protocol-explorer/all-protocols
instead of https://messari.io/protocol-explorer/all-protocols?view=tvl&page=2
when you use the TVL view, we show all of the protocols that we catalogue. For protocols that we don't have a subgraph for, we fetch their TVL from defillama
Vincent 7:06 PM
that being said, the TVL change still shouldn't show up as 0%. I've filed a bug on this and we will look into it. Thanks for the heads up!
Vincent is from Messari
Message Vincent
Home
DMs
Mentions
Search
You

Appendix:

This study is a minimum viable product (MVP) delivered in less than 24 hours for TheGraph cryptocurrency hack-a-thon hosted by the Leeds School of Business at the University of Colorado on February 25, 2023. If this study were carried to a logical conclusion, the team would recommend the following:

1. Further studies would use more robust sentence level NLP to confirm rudimentary Word Cloud assumptions. Although the Word Cloud findings are more than statistically significant (Z-Scores > 5.6, N>3000, thrice), sentence level vector comparisons using LLM technology would conclusively confirm or invalidate the findings using the latest methods.

2. Further studies would plot all 35,000+ TheGraph Subgraphs, not just 50 from the Exchanges section (here https://subgraphs.messari.io/). Adding the portfolio of Messari Subgraphs would make the plot more impressive (double y axis amplitude). Adding TheGraph Subgraphs may yield other insights.

3. This approach would be improved by systematic analysis of all 35,000+ Subgraphs hosted at TheGraph.com (https://thegraph.com/hosted-service). Automated population of graph names would feed analysis functions in perpetuity, contribute to annual reporting, and aid troubleshooting of front end software. (Original anomaly detection attempt abandoned due to hashed subgraphs/id infrastructure. Code included.)

4. NLP method of analysis may be extended, yielding more granular results, with additional corpora and Subgraphs. Further testing might yield granular analysis capable of validating outcomes agreed upon by disparate development groups within TheGraph ecosystem.

Special thank you to Professor Kai Larsen, representatives of TheGraph and Messari. Without your gracious “Subject Matter Expertise” and “Voice of the Customer” this study would not have been possible. We thank you for the opportunity.