Factors that Contribute to Crowdsourcing Success: the case of Waze

Presented by:

Giovanna Rojas
311046622

Supervisor: Prof. Joseph G. Davis

MITM Candidate
Knowledge Discovery and Management Research Group

FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

THE UNIVERSITY OF SYDNEY

Research sponsored by: Australia Awards

THE UNIVERSITY OF SYDNEY

Agenda

- Motivation and Background
- Waze
- Research Design & Solving Method
- Data Analysis and Results
- Discussion & Conclusion
- Future Work

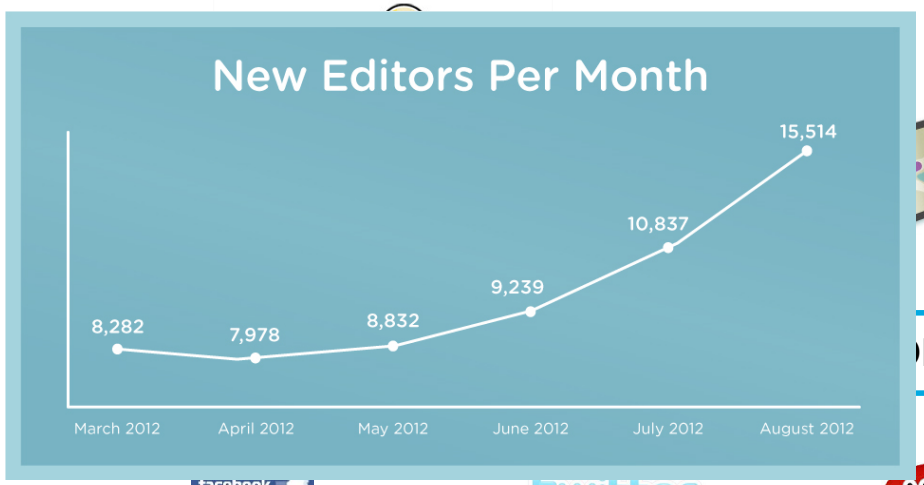
Motivation and Background

- Crowdsourcing (Howe, 2006) on the rise
- Measures of Success
  - Information Systems (DeLone & McLean, 2003)
  - FLOSS (Mockus et al, 2002; Crowston, Howison, et al., 2006, Au, Carpenter et al, 2009)
  - Crowdsourcing
    - Wikipedia's Quality (Stvilia et al, 2005) and Participation (Neus, 2001)
    - Map Tracing Quality (Haklay et al, 2010)
    - Business Perspective (Sharma, 2010)
- Contribution:
  - Fill the gap in literature regarding factors that contribute to crowdsourcing success

    "*It is not possible to improve a process without knowing what needs to be improved*"

Waze

› Volunteered Geographic Information (Goodchild, 2007)



- ✓ 36 Million Users
- ✓ 110 Countries
- ✓ 90M User reports
- ✓ 500M Map edits (Shu, 2013)
- ✓ $1.3B Google (Moskvitch, 2013)

RY SENSING

Images source: Waze website

Waze – Unlock/Update Request

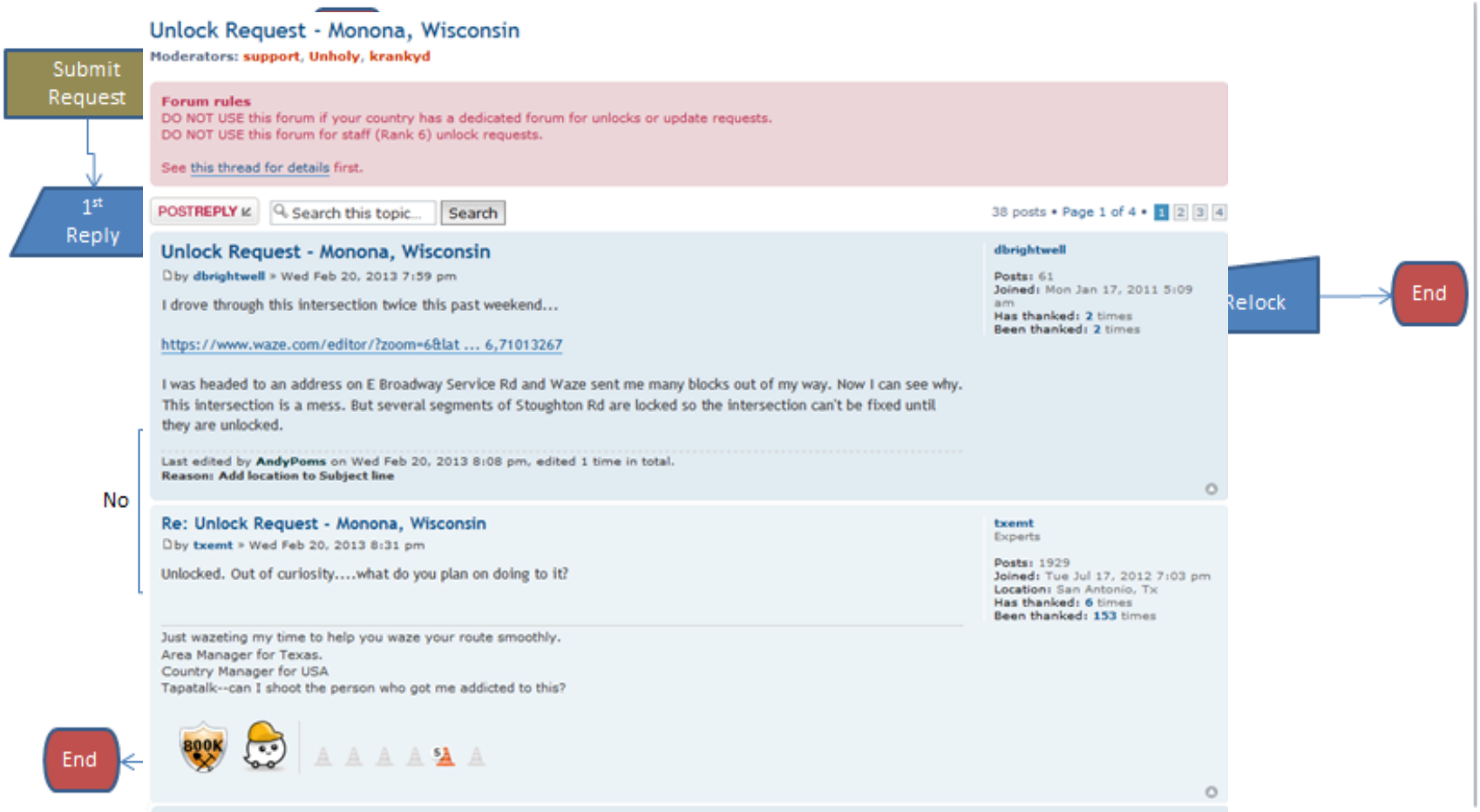
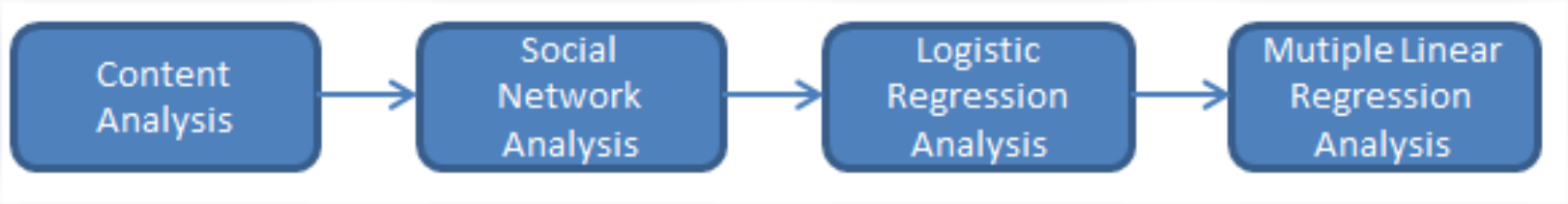


Fig. 1. Unlock Request Process

Fig. 2. Unlock Update Request Example

Research Design & Solving Method

- As of April 29th, 2013: **5,919** unlock/update requests
- **95** Sample Posts
  - Margin of error: 10% (Kotrlik et al, 2001)
  - Most popular threads,
  - Most view threads,
  - and random threads with at least 5 replies



Content & Social Network Analysis

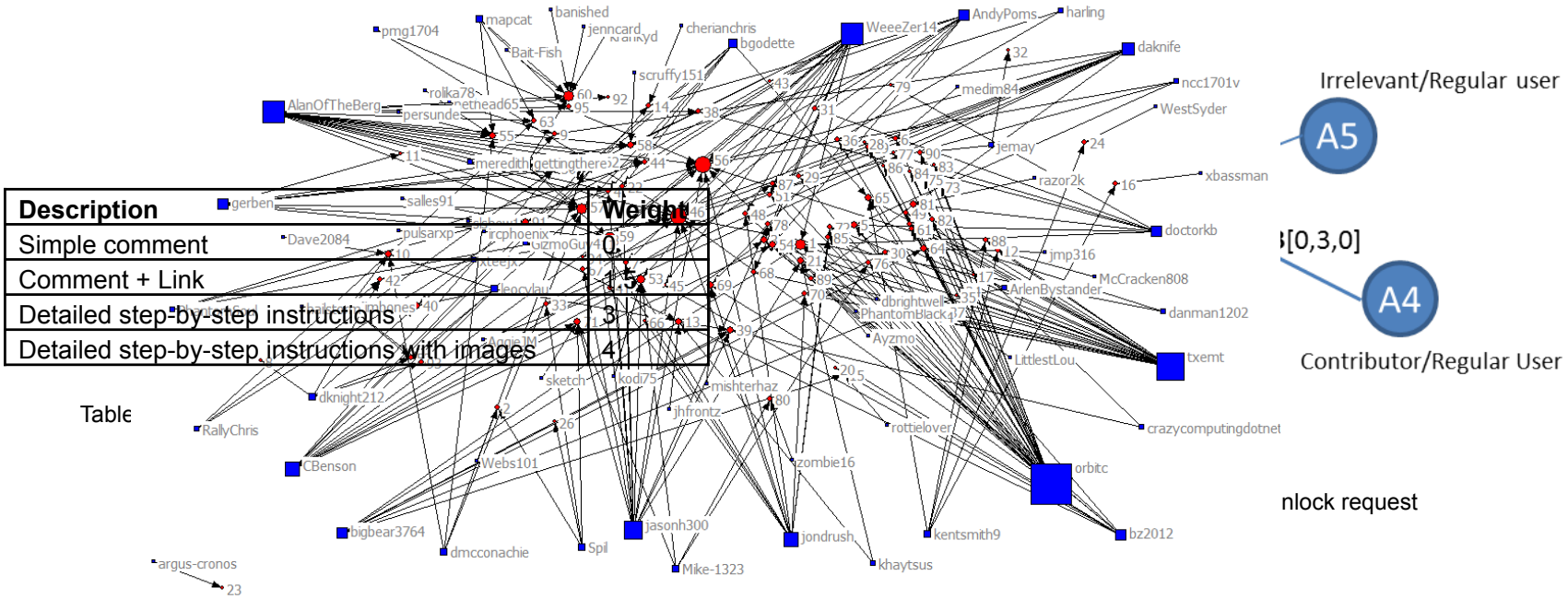


Figure 4: Waze Unlock/Update Request Social Network (Based on a sample of 95 posts). Network centralization index = 79.63%

	RT 1st Reply	Leader Centrality	Subnet Centrality	KS	Contributors	RT Finished
Mean	4:06:25	23.17	1.82	4.54	2.95	129:22:21
Median	0:43:00	24.13	1.22	1.50	2.00	14:33:00
Std Dev	9:39:08	19.02	1.34	8.16	2.22	323:16:40
Min	0:01:00	0.20	0.61	0.00	1.00	0:06:00
Max	81:33:00	57.37	7.32	52.50	12.00	1668:50:00

Table 2 – Overall Statistics for 95 posts. 63 were successfully completed

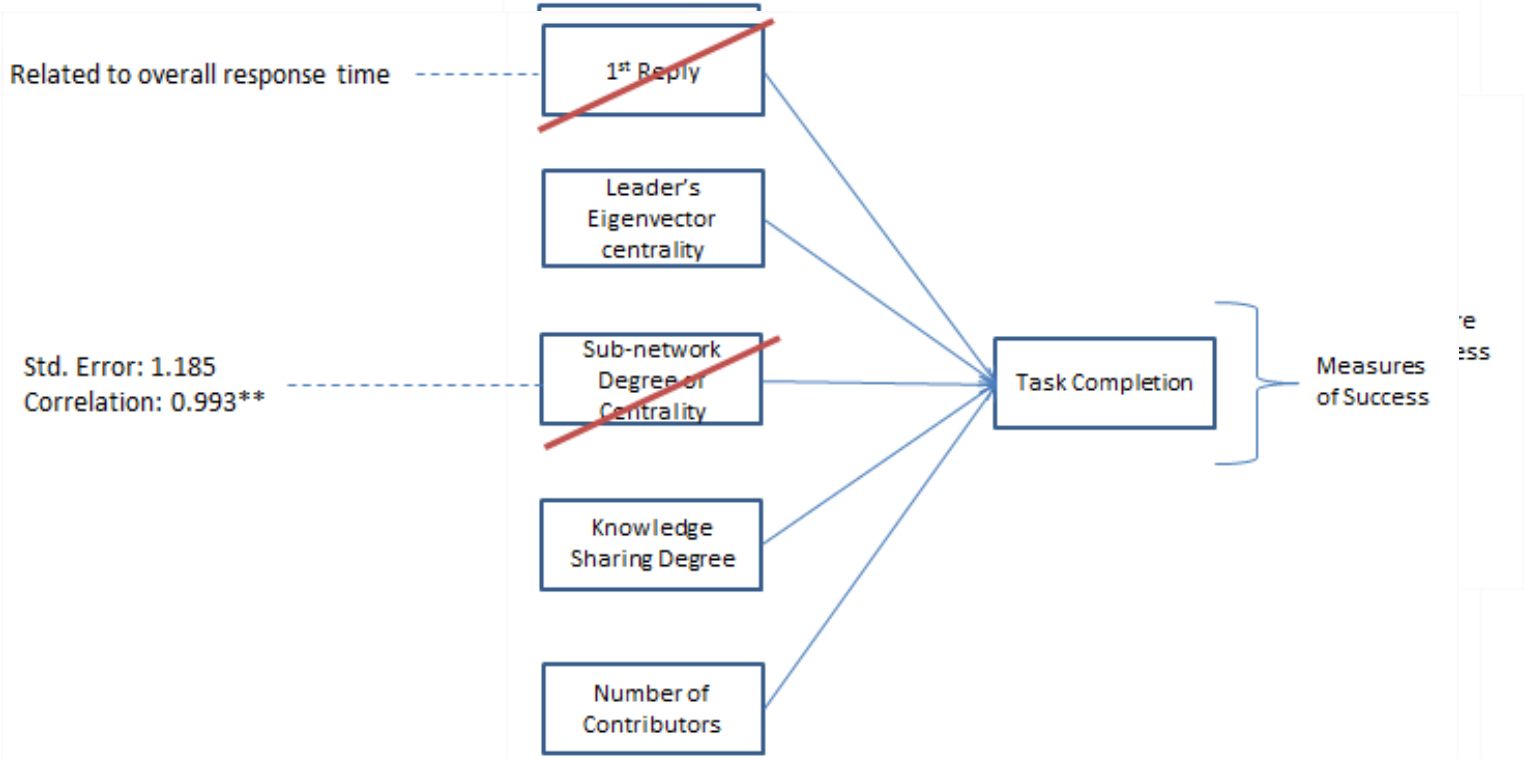
Modelling

Logistic Regression Model

$$probability\ (task\ completion) = \frac{e^{\beta_0 + \beta_1(FR) + \beta_2(LC) + \beta_3(CS) + \beta_4(KS) + \beta_5(CC)}}{1 + e^{\beta_0 + \beta_1(FR) + \beta_2(LC) + \beta_3(CS) + \beta_4(KS) + \beta_5(CC)}}$$

Multiple Linear Regression Model

$$Performance\ (task\ completion) = \beta_0 + \beta_1(FR) + \beta_2(LC) + \beta_3(CS) + \beta_4(KS) + \beta_5(CC)$$



Statistical Analysis Results

$$probability(task\ completion) = \frac{e^{\beta_0 + \beta_1(LC) + \beta_2(KS) + \beta_3(CC)}}{1 + e^{\beta_0 + \beta_1(LC) + \beta_2(KS) + \beta_3(CC)}}$$

	Chi-square	df	Sig.
Step 1 Model	7.054	3	**.070**

Table 3 – Summary of the logistic model to predict task completion

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	LeaderCentrality	.000	.000	.217	1	.642	1.000	1.000	1.000
	KS	.074	.044	2.812	1	**.094**	1.077	.988	1.174
	Contributors	-.325	.152	4.532	1	**.033**	.723	.536	.975
	Constant	1.416	.424	11.179	1	.001	4.122		

Table 4 – Variables' coefficients and odds Ratio for the logistic model to predict task completion

Statistical Analysis Results (cont.)

$$performance\ (task\ completion) = \beta_0 + \beta_1(LC) + \beta_2(KS) + \beta_3(CC)$$

R	R Square	Std. Error of the Estimate
.527[a]	.277	283.95943

Table 5 – Summary of the linear regression model for task completion performance

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-73.167	78.878		-.928	.357
	LeaderCentrality	-.624	1.892	-.037	-.330	.743
	KS	-21.640	6.395	-.632	-3.384	**.001**
	Contributors	115.118	24.940	.859	4.616	**.00002**

Table 6 – Variables' coefficients and significance levels of the linear regression model for task completion performance

Evaluation performed using the 63 successful threads (out of 95)

Discussion and Conclusion

- Knowledge sharing
  - Informal learning [tacit knowledge] (Eraut, 2000)
  - Interest (bring others up-to-speed)
- Specialized tasks
- Avg. number of contributors: ~3 (Mockus et al, 2002; Au, Carpenter, et al, 2009)

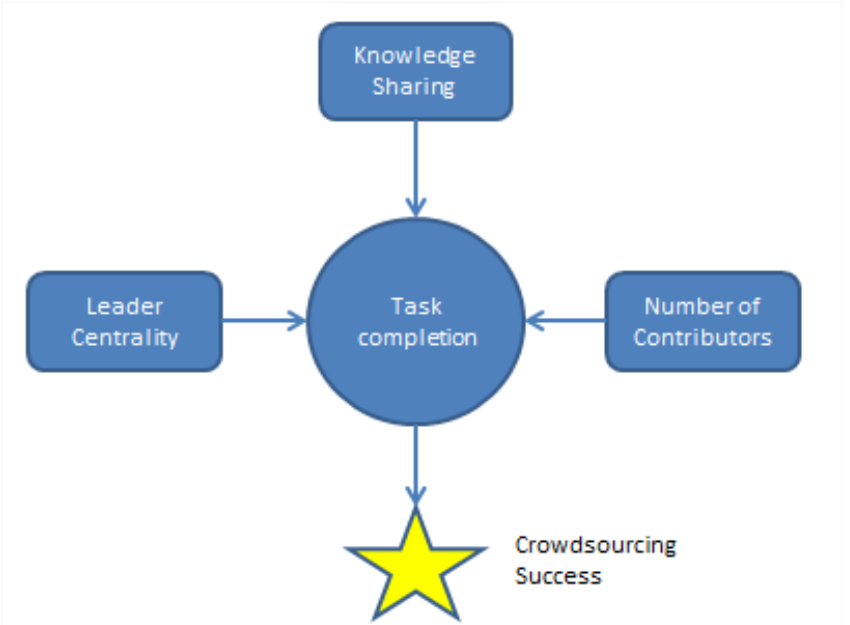


Fig. 5: Factors that contribute to crowdsourcing success. Empirical model.

Future Work

- Increase sample size
- Expand to other projects (INRIX, Google Map Maker, OpenStreetMaps)
- Improve the model (independent variables)
- Longitudinal data (knowledge creation)

THE UNIVERSITY OF SYDNEY

References

- Crowston, K., J. Howison, et al. (2006). Information systems success in Free and Open Source Software development: Theory and measures. *Software Process, Improvement and Practice,* **11**(2): 123-148.
- DeLone, W. H. and E. R. McLean (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of management information systems*, **19**(4): 9-30.
- Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology*, **70**: 113-113.
- Goodchild, M. (2007). Citizens as sensors: the world of volunteered geography. *Geo. Journal*, **69**(4): 211-221.
- Haklay, et al. (2010). How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *CARTOGR J*, **47**(4): 315-322.
- Howe, J. (2006). The rise of crowdsourcing. Wired magazine, **14**(6): 1-4.
- Kameshwaran, S., S. Mehta, et al. (2009). Analyses for Service Interaction Networks with Applications to Service Delivery. *SDM, SIAM.*
- Kotrlik, J. W. and C. C. Higgins (2001). Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research. *Information technology, learning, and performance journal*, **19**(1): 43.
- Latane, B., K. Williams, et al. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, **37**(6): 822-832.
- Mockus, A., R. T. Fielding, et al. (2002). Two case studies of open source software development: Apache and Mozilla. *ACM Trans. Softw. Eng. Methodol.*, **11**(3): 309-346.
- Moskvitch, K. (2013, June 11). *Google seals Waze app sat-nav takeover.* Retrieved June 11, from http://www.bbc.co.uk/news/technology-22864712

References (cont.)

- Neus, A. (2001). Managing Information Quality in Virtual Communities of Practice. *IQ 2001: The 6th International Conference on Information Quality at MIT.*
- Raisanen, T. and H. Oinas-Kukkonen (2008). A System Architecture for the 7C Knowledge Environment. *Proceedings of the 2008 conference on Information Modelling and Knowledge Bases XIX*, IOS Press**:** 217-236.
- Shu, C. (2013, Feb. 6). *Nav App Waze Says 36M Users Shared 900M Reports, While 65K Users Made 500M Map Edits.* Retrieved Feb 6, from http://techcrunch.com/2013/02/06/nav-app-waze-says-36m-users-shared-900m-reports-while-65k-users-made-500m-map-edits/
- Waze Website (2013) [Online] http://www.waze.com

THE UNIVERSITY OF SYDNEY

› Thank you

Q&A