



从连接中挖掘数据价值

华扬联众研发副总裁 李响



从连接中挖掘数据价值

华扬联众研发副总裁 李响



Contents

- 1 连接的现状和发展
- 2 对营销的挑战
- 3 技术方案
- 4 应用场景

Contents

- 1 连接的现状和发展
- 2 对营销的挑战
- 3 技术方案
- 4 应用场景



1

Part

连接的现状和发展



月活用户数10亿，
人均好友数超过100



每天32亿“like”（基于
2012数字）



每日交易2100万笔
(基于2017数字)



2016年新增会员数
7000万

各类连接关系遍布于现代社会





1

Part

连接的现状和发展



月活用户数10亿，
人均好友数超过100



每天32亿“like”（基于
2012数字）



每日交易2100万笔
(基于2017数字)



2016年新增会员数
7000万

各类连接关系遍布于现代社会

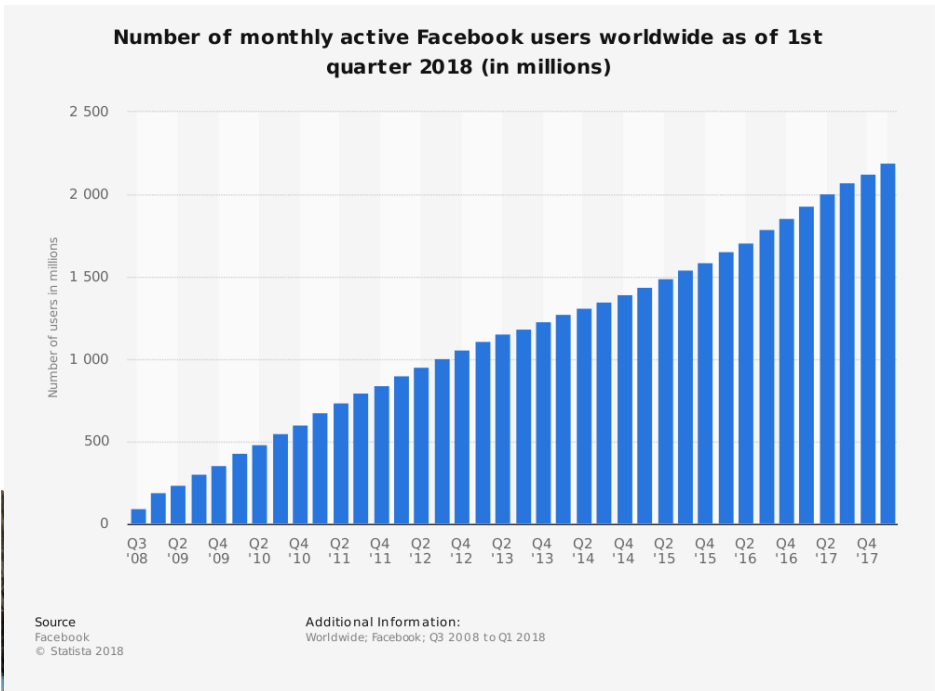




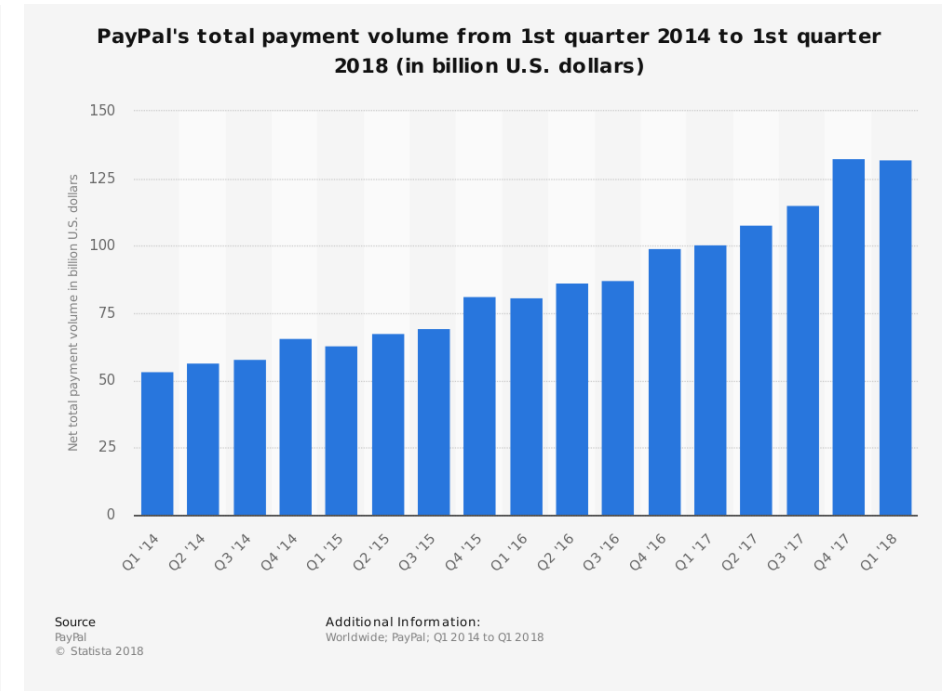


连接的现状和发展

Fb全球月活用户数



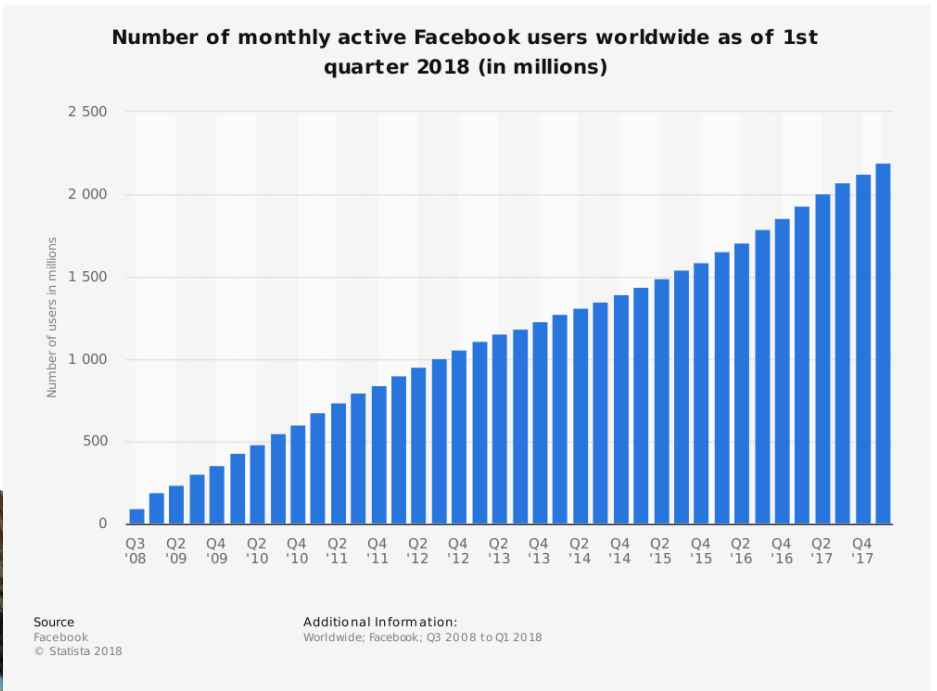
第三方支付的季节交易量



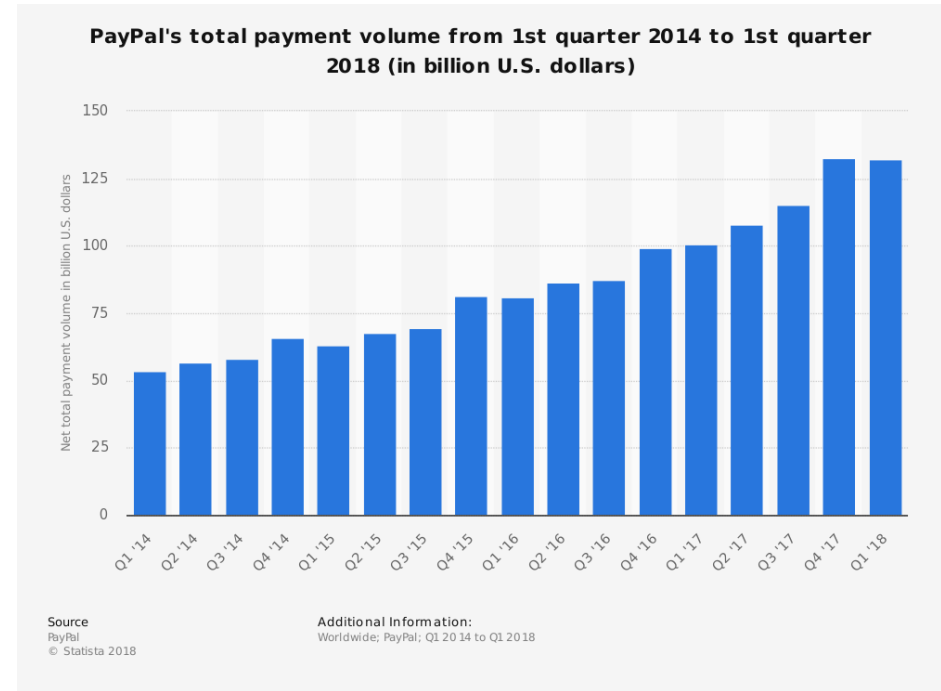
From Statista 2018

连接的现状和发展

Fb全球月活用户数



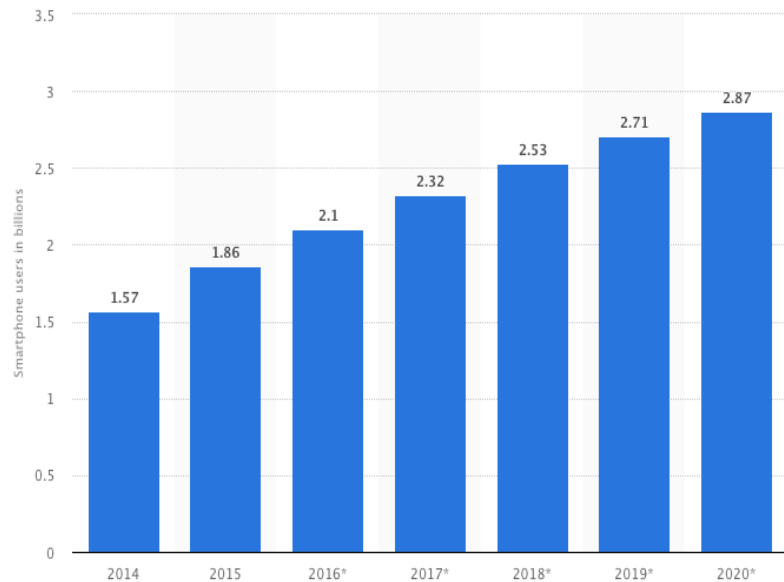
第三方支付的季节交易量



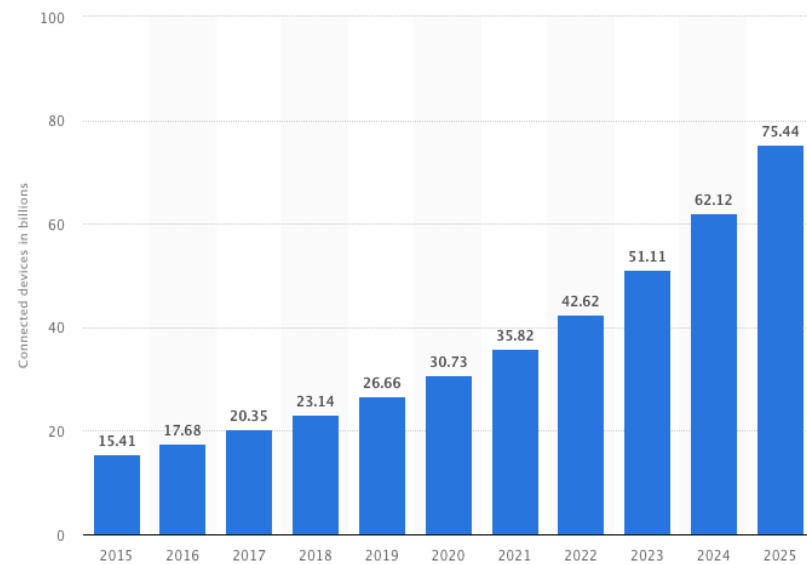
From Statista 2018

连接的现状和发展

全球智能手机用户数



全球IoT设备数量

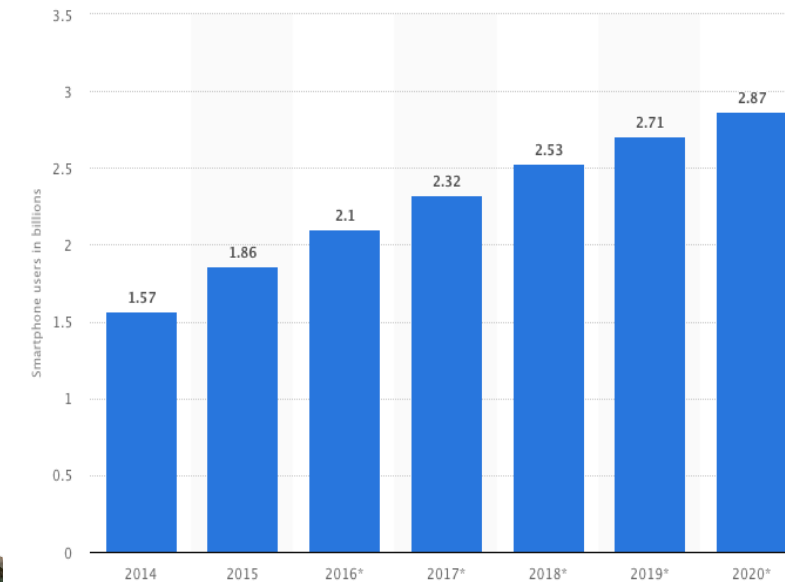


软硬件方面的发展会引入更多的连接关系

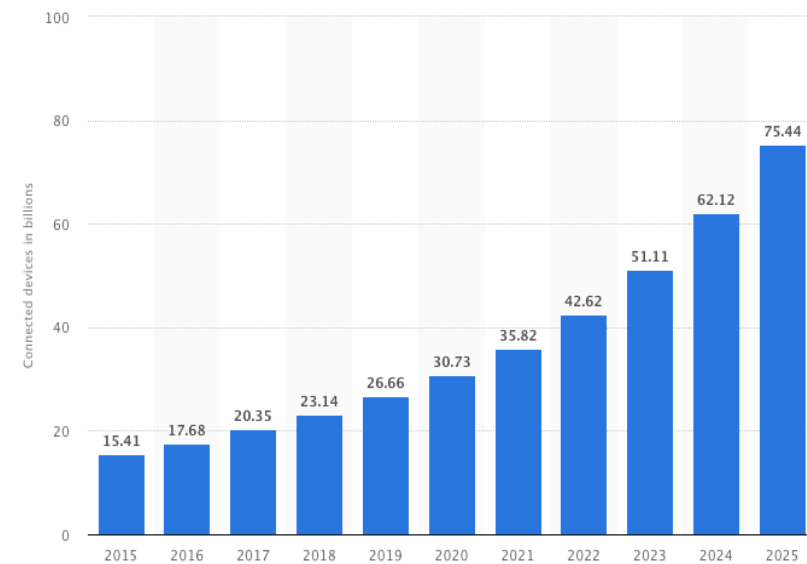
From Statista 2018

连接的现状和发展

全球智能手机用户数



全球IoT设备数量



软硬件方面的发展会引入更多的连接关系

From Statista 2018



Part Two

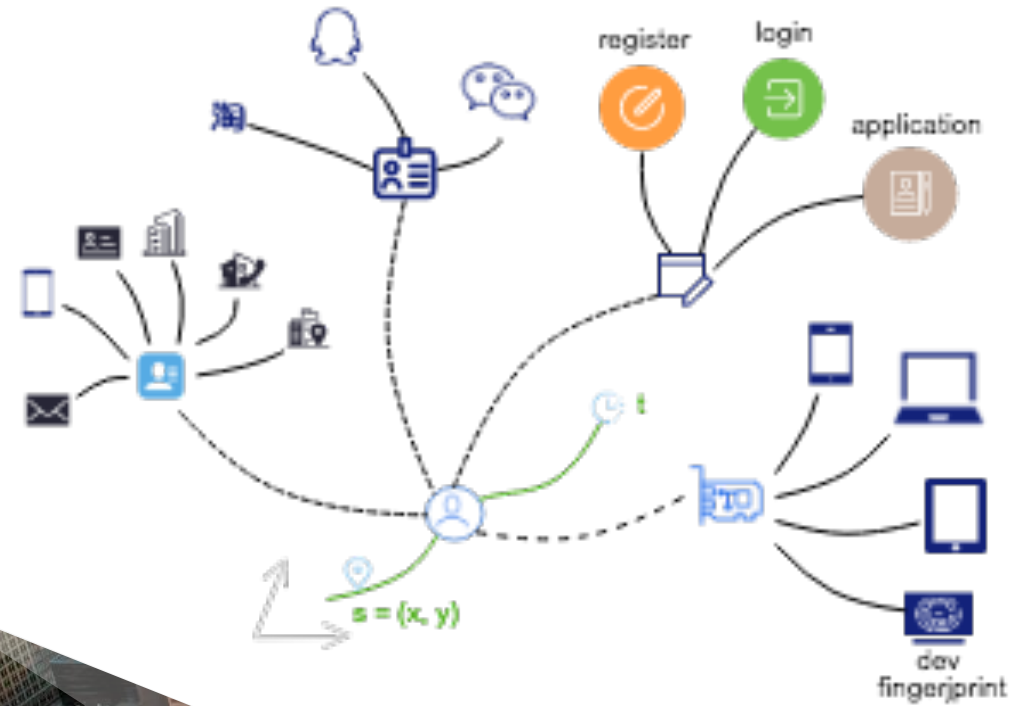
对营销的挑战



Part Two

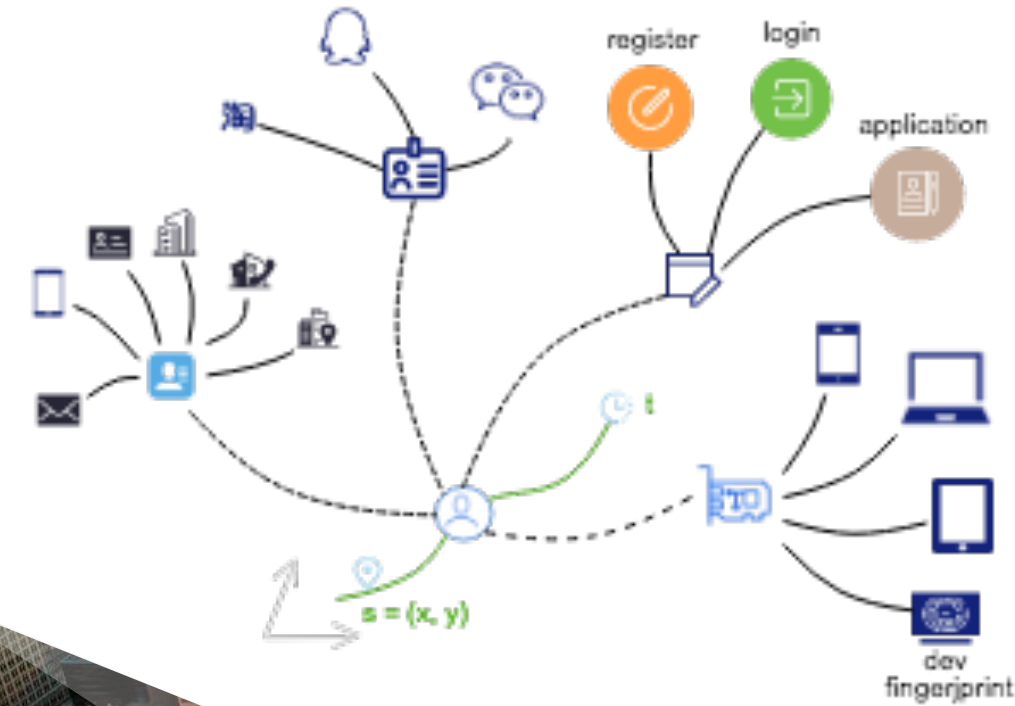
对营销的挑战





entities			
	account	geospatial: s	
	event		
	person		
	device		

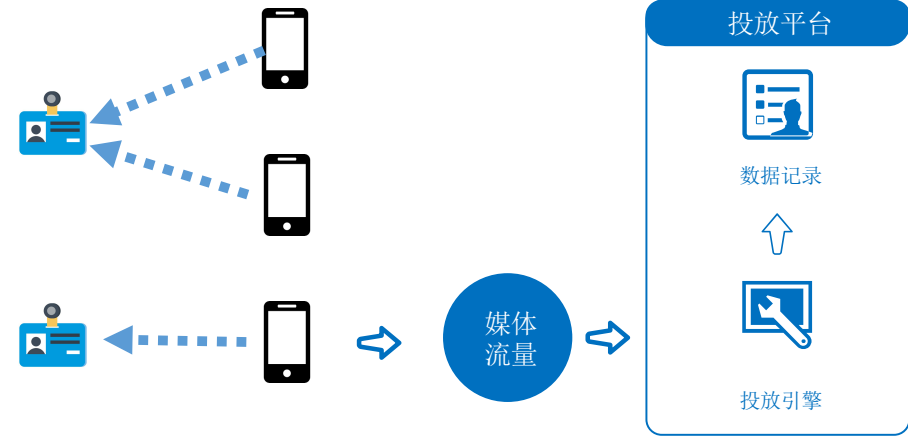
连接网络中不可避免含有不同类型的实体 (Vertex)



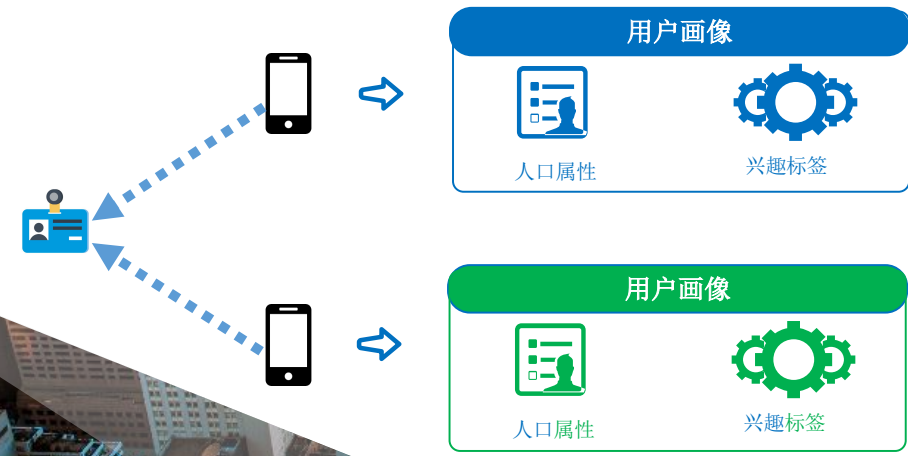
entities			
	account	 geospatial: s	 temporal: t
	event		
	person		
	device		

连接网络中不可避免含有不同类型的实体 (Vertex)

对营销的挑战

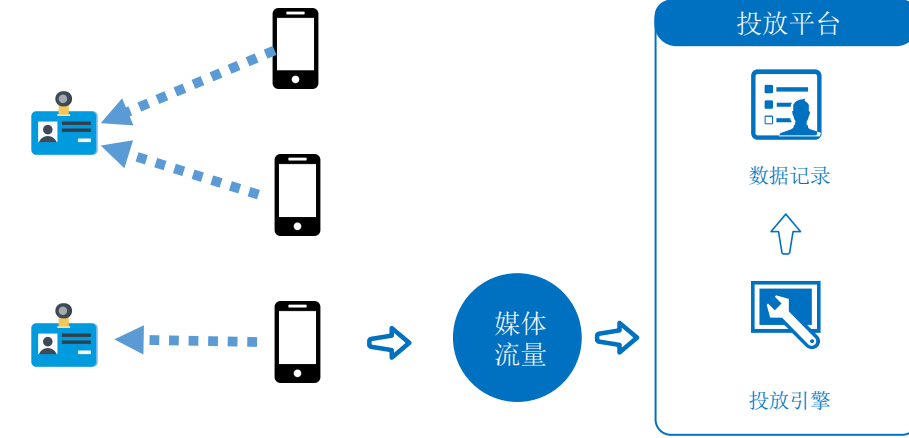


对以个人用户为主要对象的精准投放的控制和评估造成困难

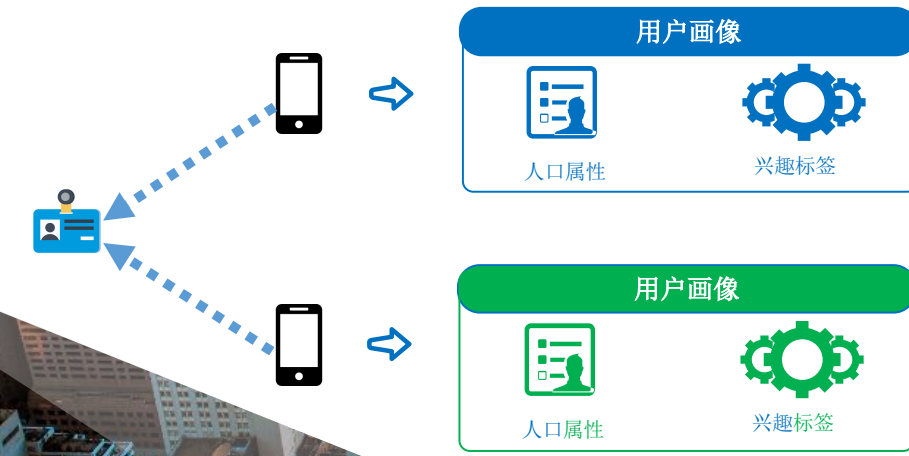


用户画像信息的碎片化

对营销的挑战

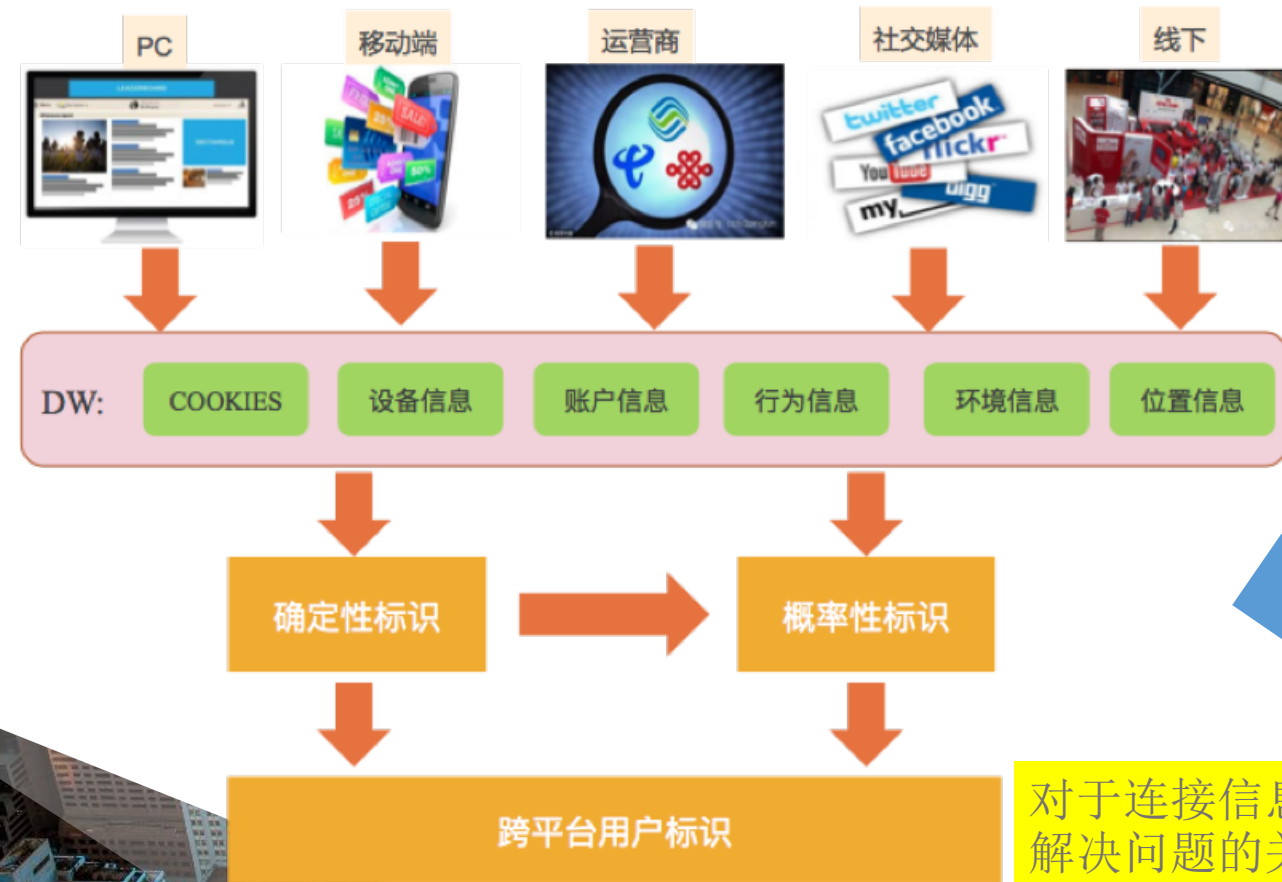


对以个人用户为主要对象的精准投放的控制和评估造成困难

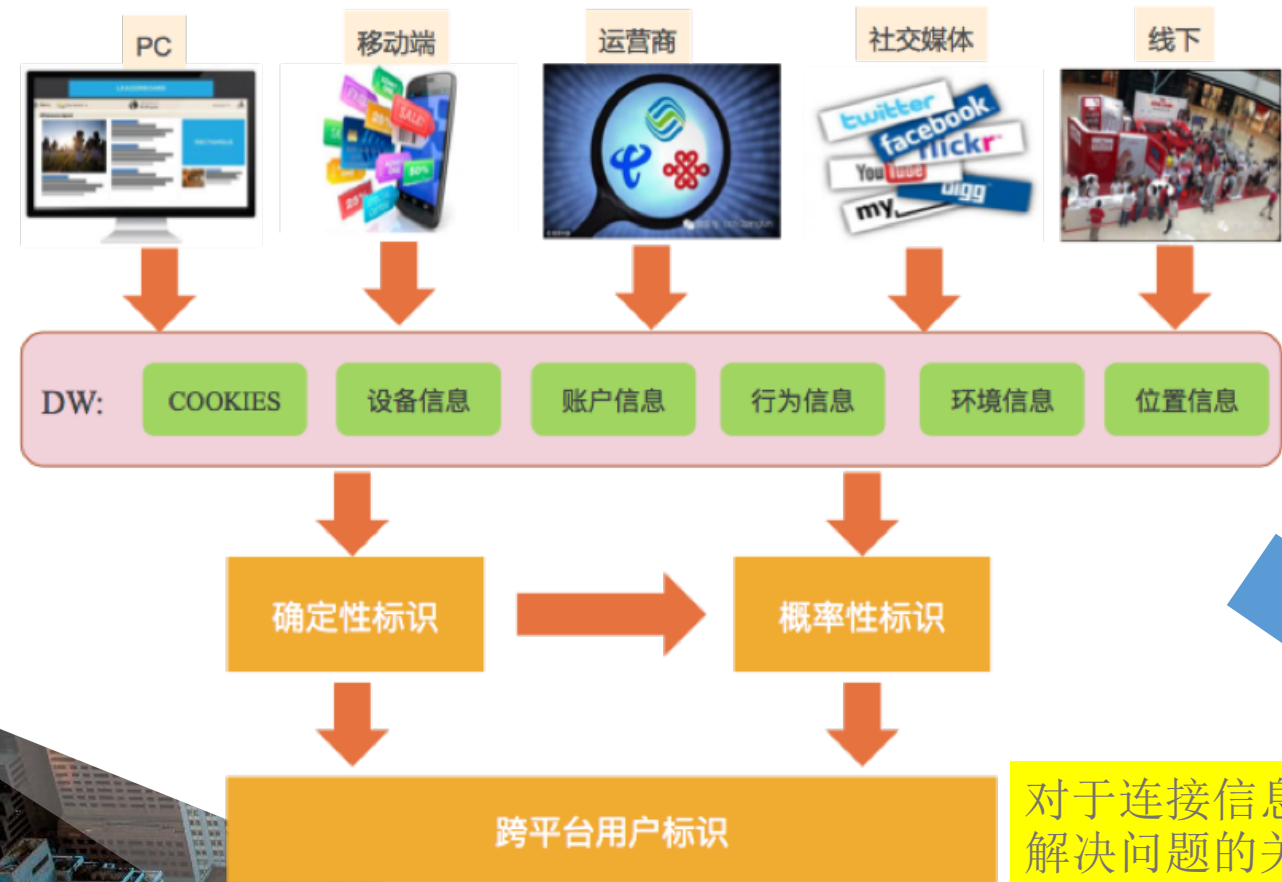


用户画像信息的碎片化

对营销的挑战



对营销的挑战





Part Three

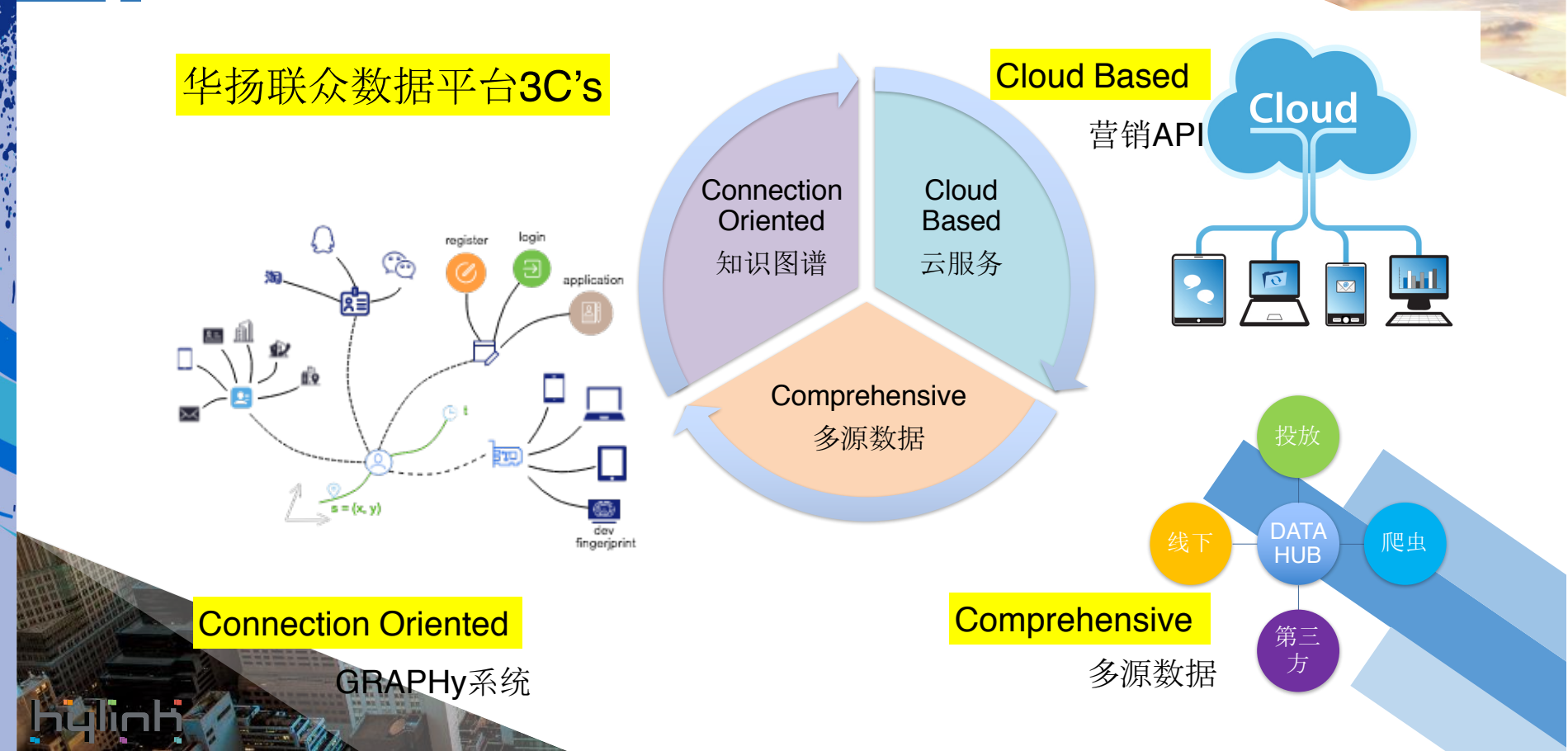
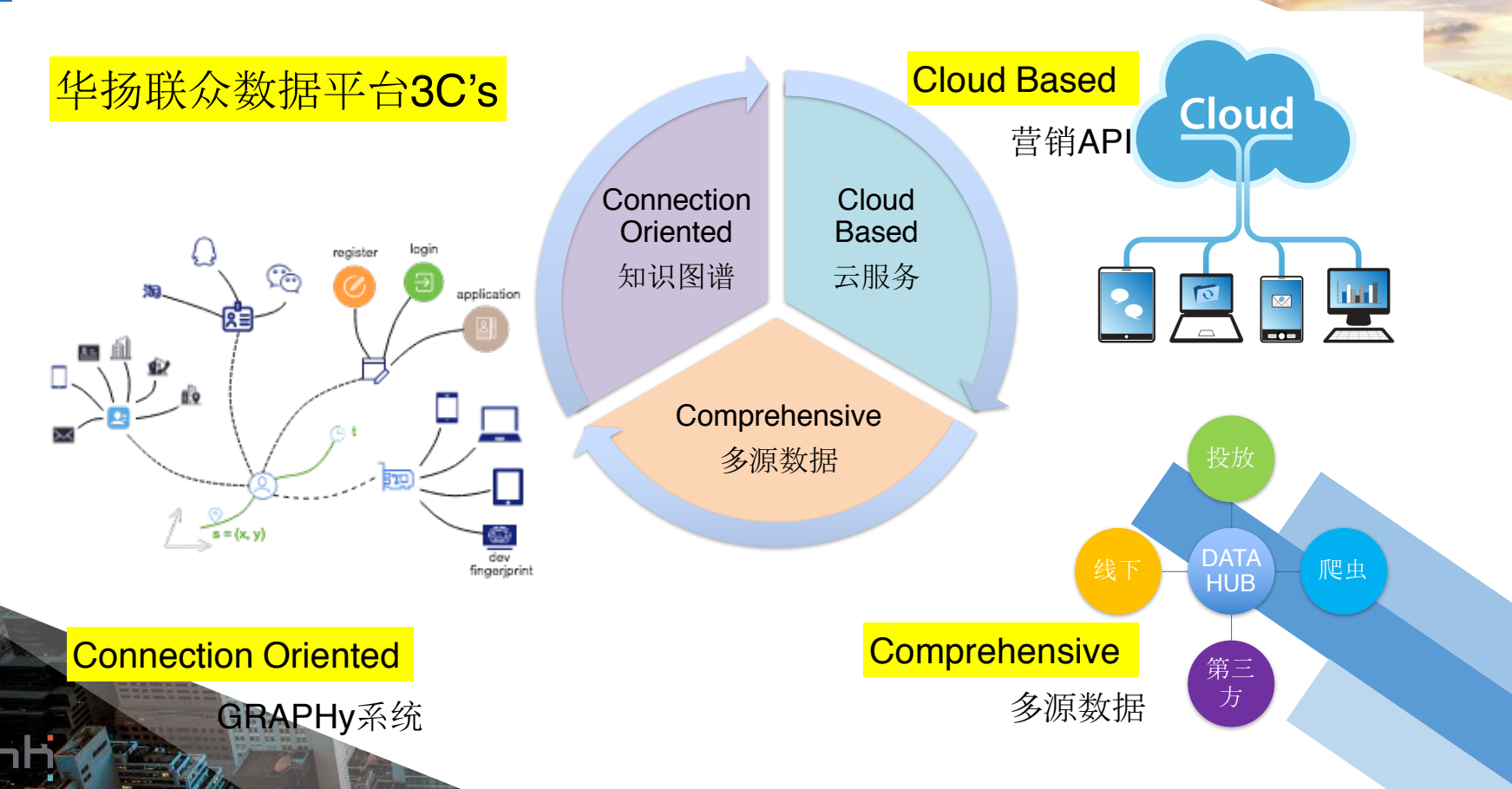
技术方案

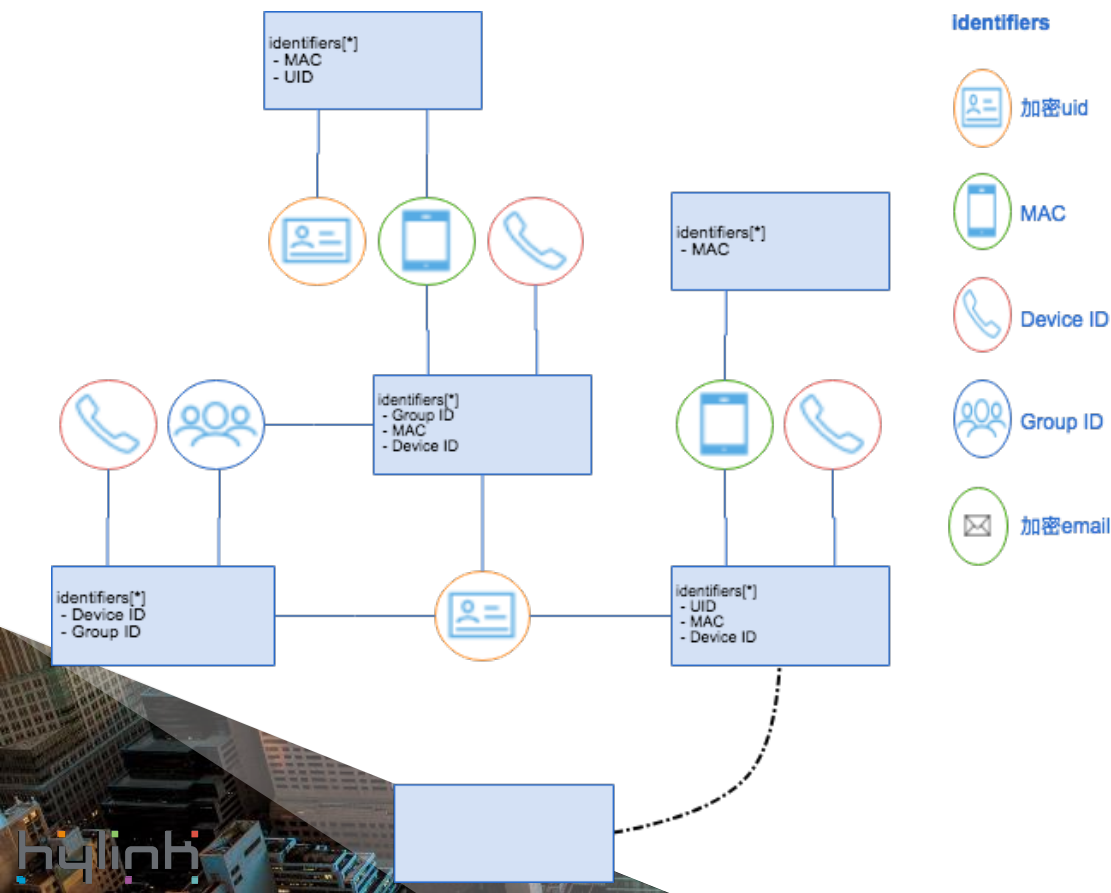


Part Three

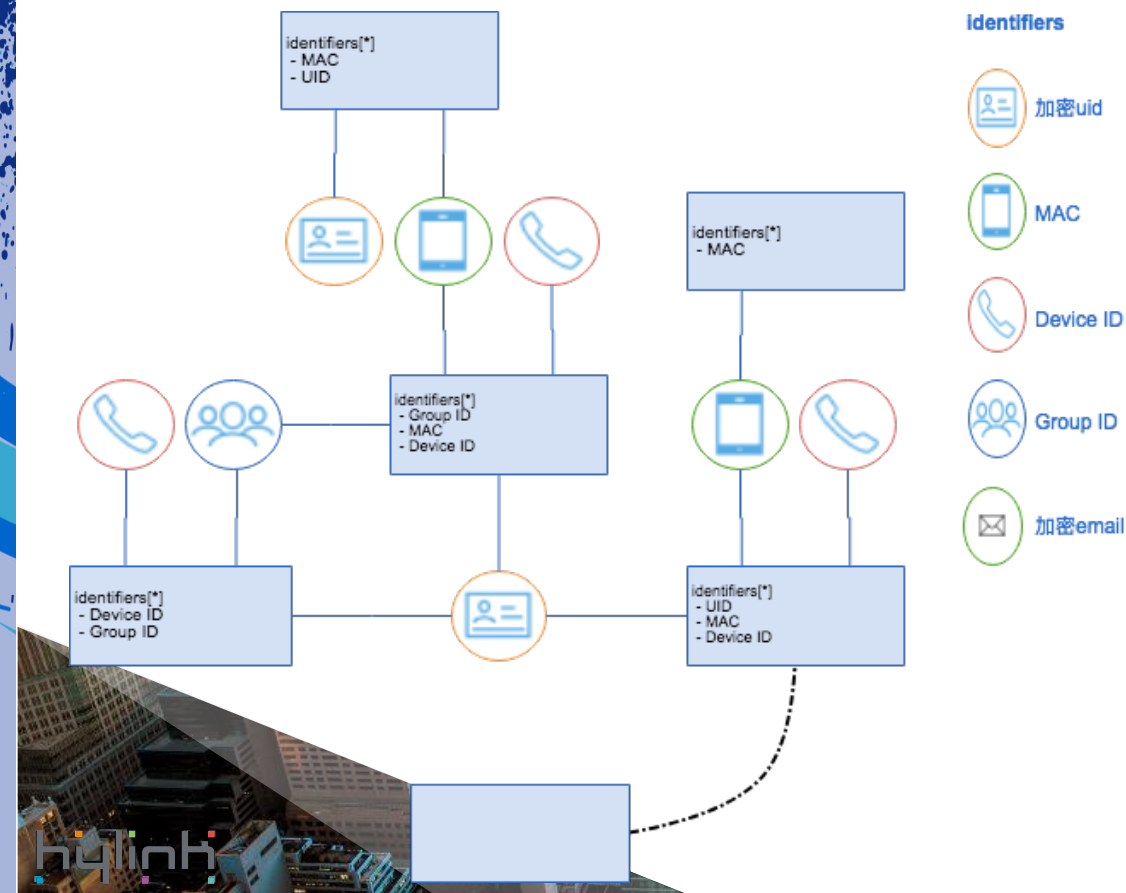
技术方案



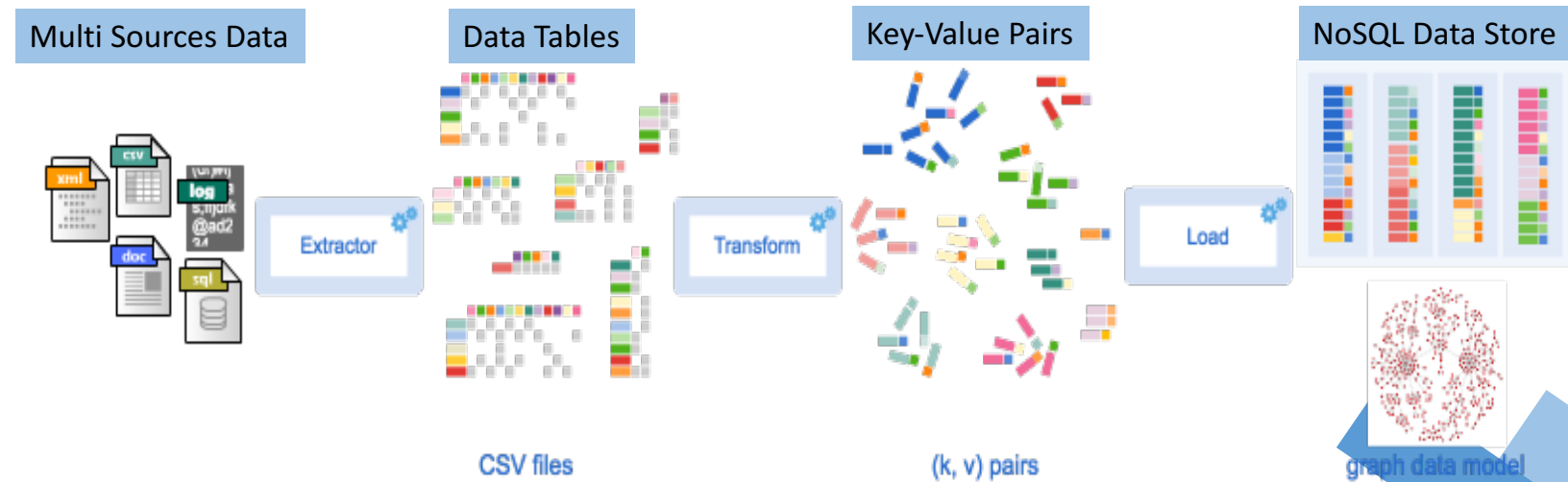




- 节点：实体
- 边：相同或类似的属性，事件
- 支持超大规模图网络的存储和查询
- 高性能
- 可扩展
- 高容错

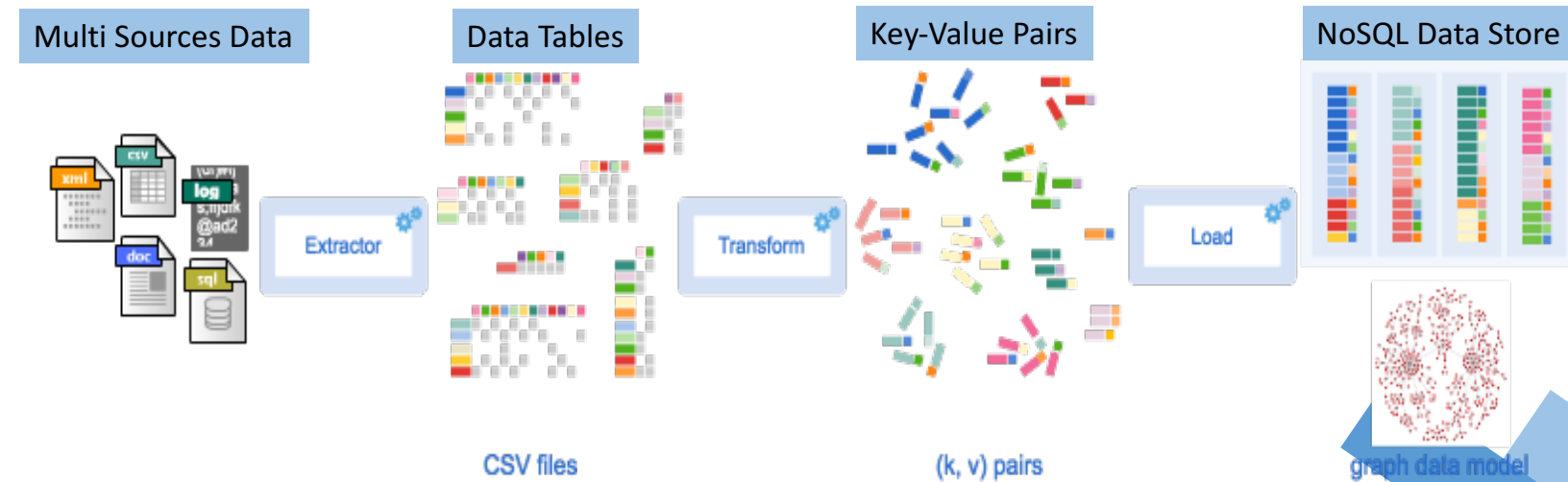


- 节点：实体
- 边：相同或类似的属性，事件
- 支持超大规模图网络的存储和查询
- 高性能
- 可扩展
- 高容错



Graph Repository of Analytics Platform for Hylink:

- 原始数据处理
- 分布式NoSQL数据库存储



Graph Repository of Analytics Platform for Hylink:

- 原始数据处理
- 分布式NoSQL数据库存储



Part Four

应用场景

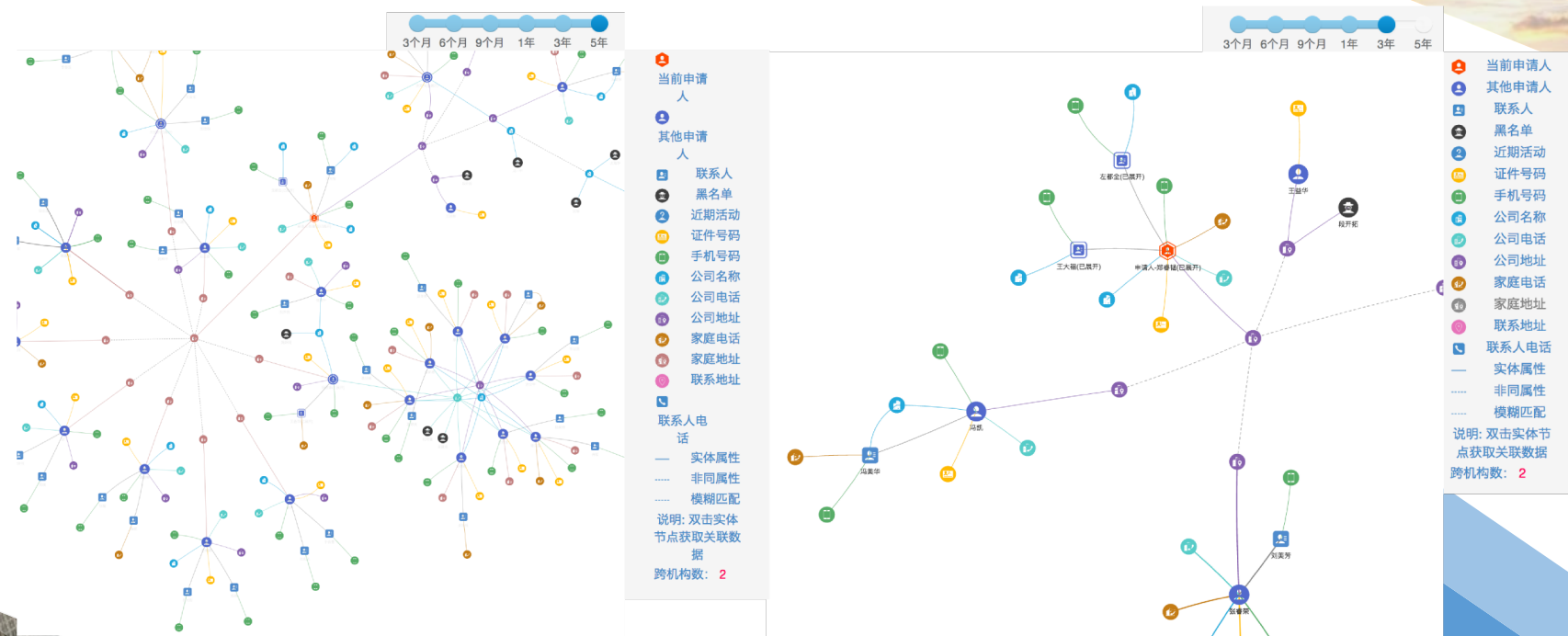


Part Four

应用场景

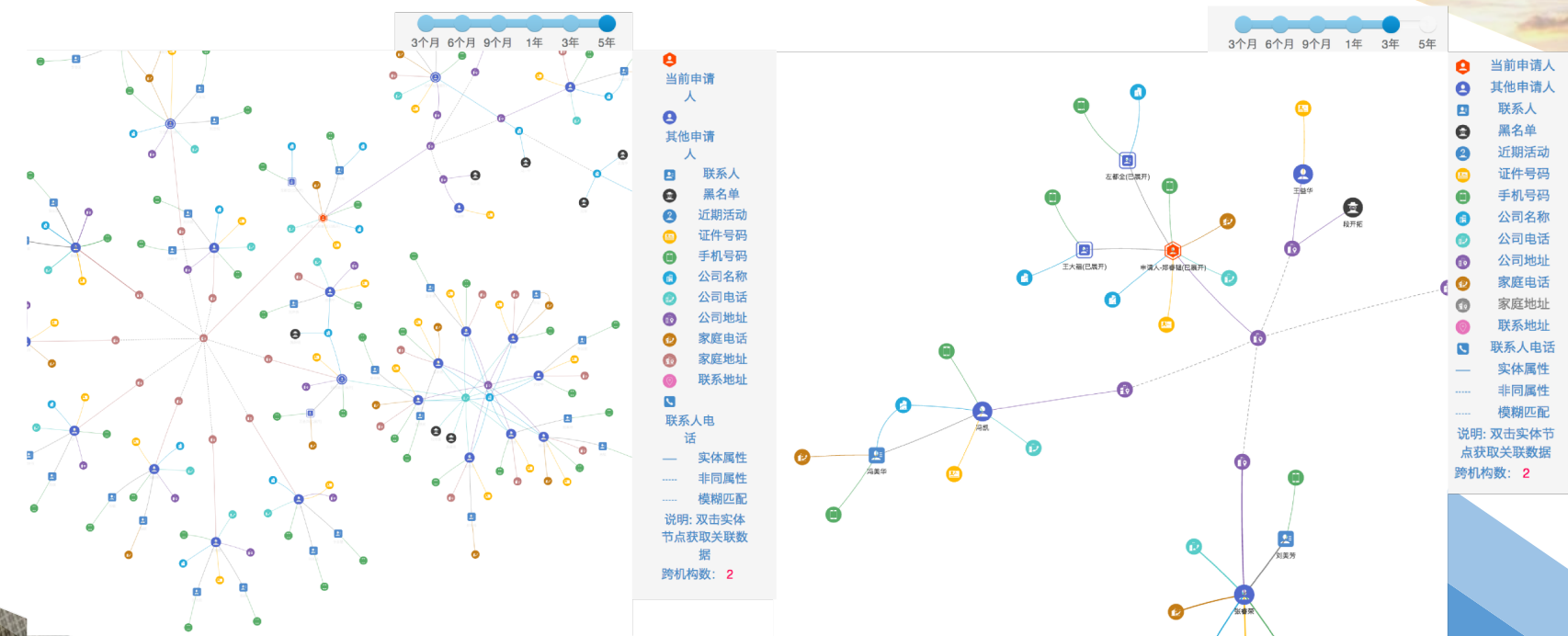


应用场景：关联可视化

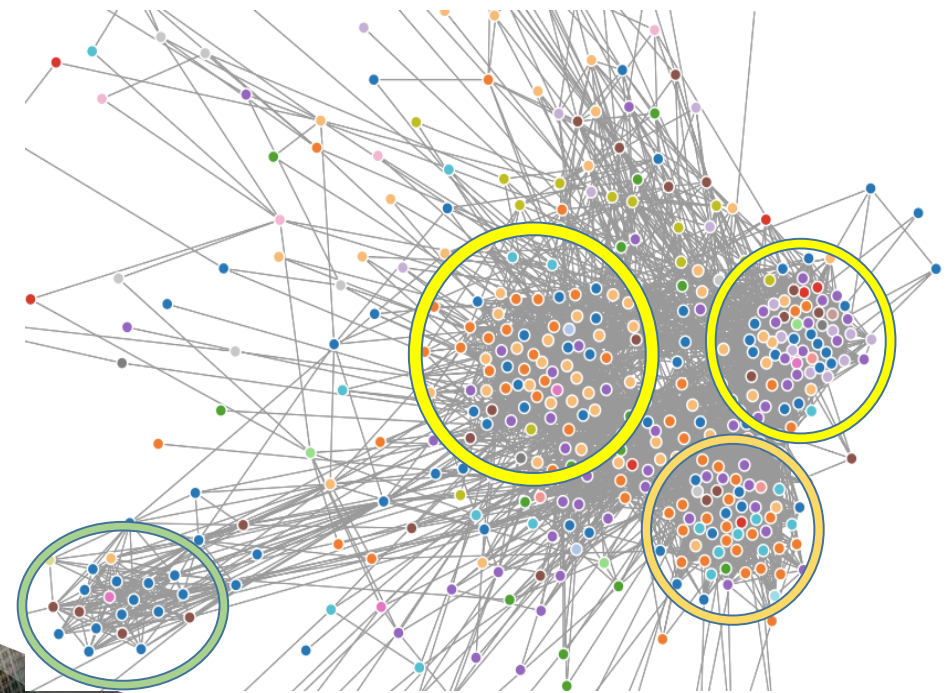


利用GRAPHy系统对节点的关联关系进行实时展示

应用场景：关联可视化



利用GRAPHy系统对节点的关联关系进行实时展示



常见相似性度量

▣ 闵可夫斯基Minkowski距离

$$\text{dist}(X, Y) = \left(\sum_{i=1}^n (x_i - y_i)^p \right)^{\frac{1}{p}}$$

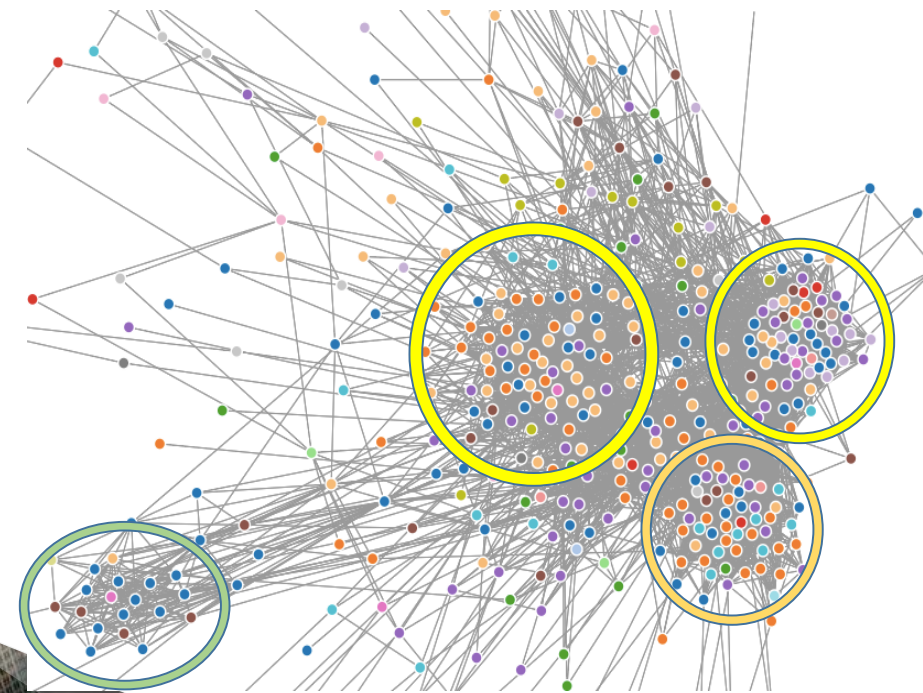
▣ 杰卡德相似系数 (Jaccard)

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

▣ 余弦相似度 (cosine similarity)

$$\cos(\theta) = \frac{a^T b}{|a| \cdot |b|}$$

基于GRAPHy系统应用相似性指标进行客户分群



常见相似性度量

▣ 闵可夫斯基Minkowski距离

$$\text{dist}(X, Y) = \left(\sum_{i=1}^n (x_i - y_i)^p \right)^{\frac{1}{p}}$$

▣ 杰卡德相似系数 (Jaccard)

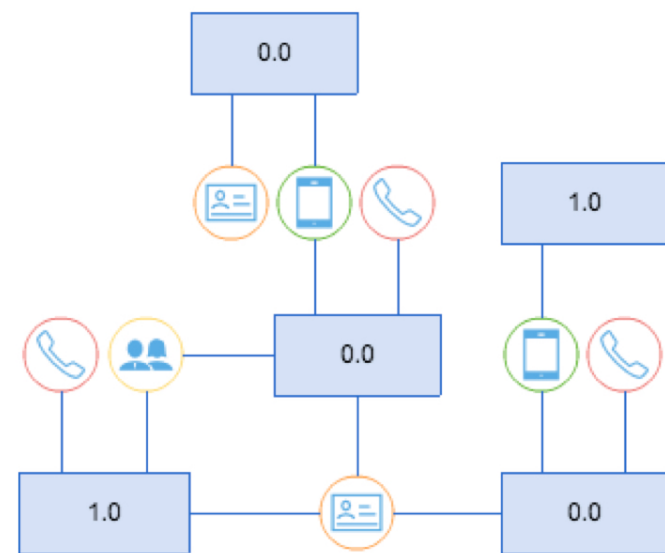
$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

▣ 余弦相似度 (cosine similarity)

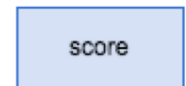
$$\cos(\theta) = \frac{a^T b}{|a| \cdot |b|}$$

基于GRAPHy系统应用相似性指标进行客户分群

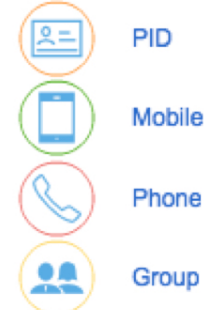
基本原型



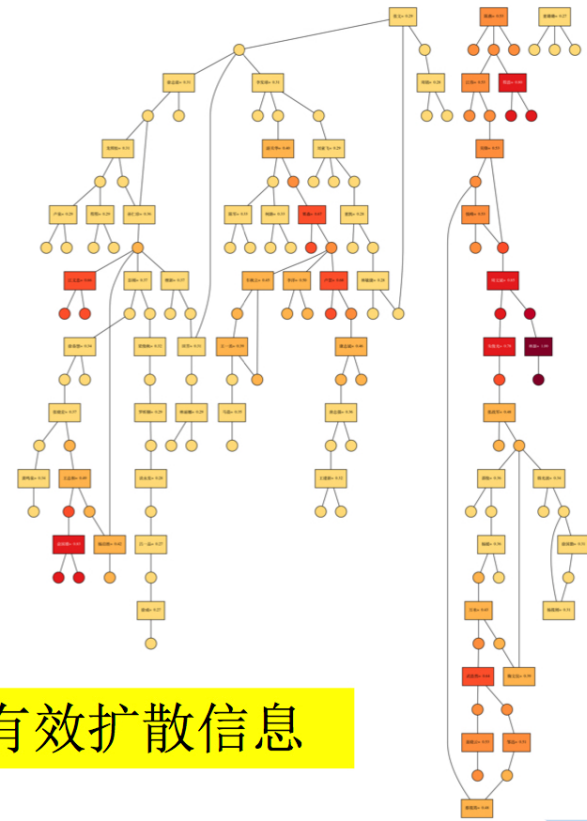
Source Nodes



Link Nodes

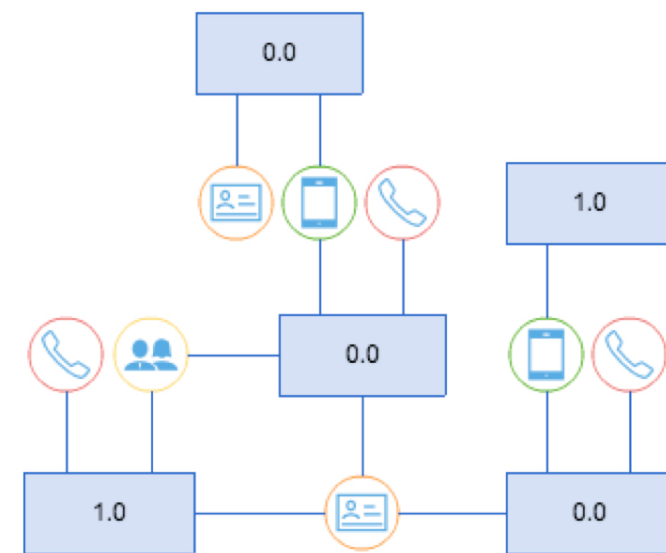


迭代计算

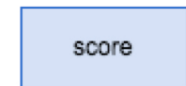


利用迭代计算在一定范围内有效扩散信息

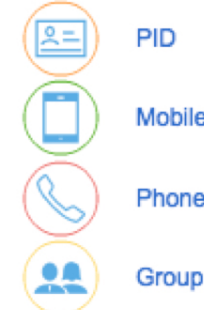
基本原型



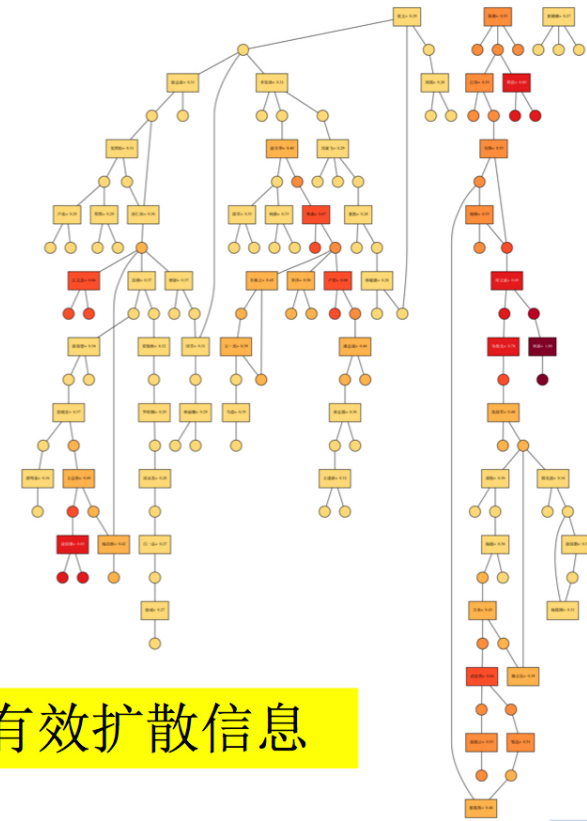
Source Nodes



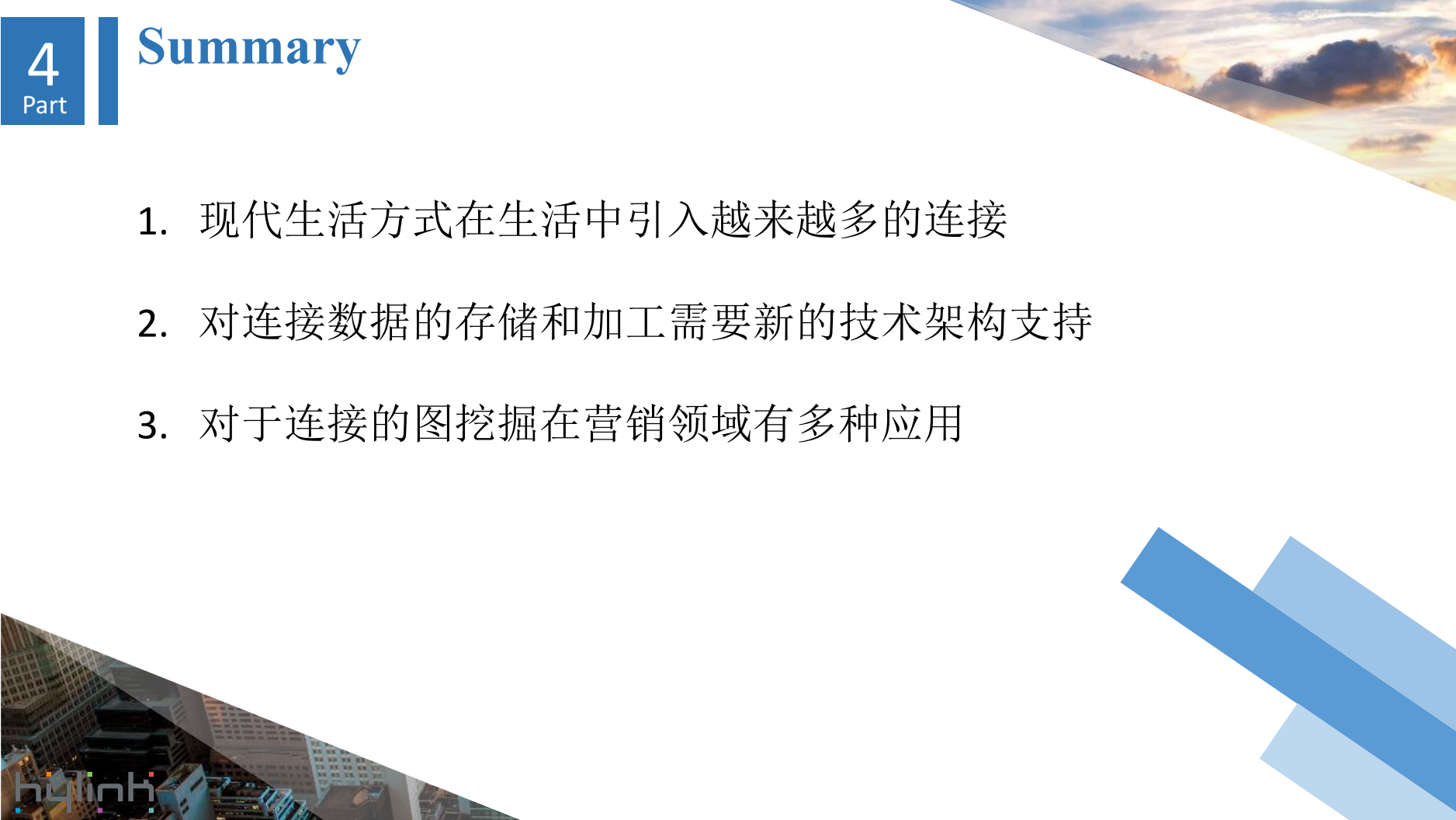
Link Nodes



迭代计算



利用迭代计算在一定范围内有效扩散信息

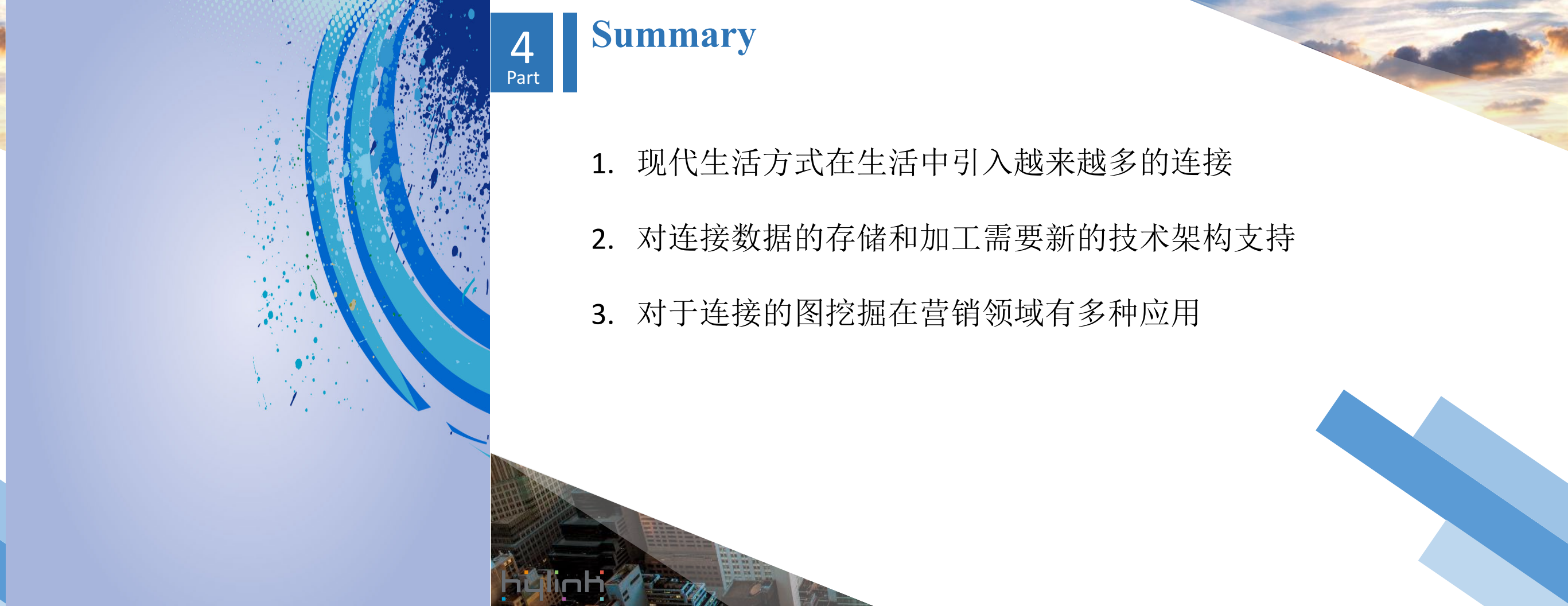


4
Part

Summary

1. 现代生活方式在生活中引入越来越多的连接
2. 对连接数据的存储和加工需要新的技术架构支持
3. 对于连接的图挖掘在营销领域有多种应用





4
Part

Summary

1. 现代生活方式在生活中引入越来越多的连接
2. 对连接数据的存储和加工需要新的技术架构支持
3. 对于连接的图挖掘在营销领域有多种应用



华扬联众Hylink
是中国领先的专注于提供互联网综合
营销服务的数字传媒公司

成立于1994年，2002年以来专注为客户提
供互联网综合营销服务，2017年8月2日上
交所上市，股票代码603825

华扬联众于2016年在美国硅谷成
立D-LAB,专注于利用大数据和机
器学习技术解决商业应用问题。



华扬联众Hylink
是中国领先的专注于提供互联网综合
营销服务的数字传媒公司

成立于1994年，2002年以来专注为客户提
供互联网综合营销服务，2017年8月2日上
交所上市，股票代码603825

华扬联众于2016年在美国硅谷成
立D-LAB,专注于利用大数据和机
器学习技术解决商业应用问题。





THANKS

2018.5.23



THANKS

2018.5.23