Below we show how the SQL queries $Q_1^p$ and $Q_2^p$ are generated for validating CFD$^p$s in $\Sigma_{CFD}^p$, which is an extension of the SQL techniques for CFDs and eCFDs discussed in [7] and [2], respectively.

The queries $Q_1^p$ and $Q_2^p$ for the violations of $\Sigma_{CFD}^p$ are given as follows, which capitalize on the data table $enc_{L}$, $enc_{R}$ and $\pi_{enc}$ that encode CFD$^p$s in $\Sigma_{CFD}^p$.

$Q_1^p$: select $R_i \ast$ from $R_i, enc_{L}, L, enc_R, R, enc_\pi \ N$

where $L.cid = R.cid$ and $R_i.X \cong L$ and $R_i.Y \cong N$ and

not $(R_i.Y \cong R$ and $R_i.X \cong N)$

$Q_2^p$: select distinct $X_L$

from (select $X_L$ as $cid$, $X_L$, $Y_R$ from $R_i, enc_L, L, enc_R, R, enc_\pi \ N$

where $L.cid = R.cid$ and $R_i.X \cong L$ and

$R_i.Y \cong N$ and $R.Y \neq \pi_{enc}$) as $M$

group by $cid$, $X_L$ having count (distinct $Y_R$) > 1

Here (1) $X = \{A_1, \ldots, A_m\}$ and $Y = \{B_1, \ldots, B_m\}$ are the sets of attributes in LHS and RHS of $\Sigma_{CFD}^p$ respectively; (2) $R_i.X \cong L$ is the conjunction of

$L.A_j$ is null or $R_i.A_j = L.A_j$ or ($L.A_j$ is not null and $R_i.A_j > L.A_j$),

and ($L.A_j$ is not null or $R_i.A_j < L.A_j$) and ($L.A_j$ is null or $R_i.A_j < L.A_j$)

and ($L.A_j$ is null or $R_i.A_j > L.A_j$) for each $j \in [1, m_1]$; (3) $R_i.Y \cong R$ is defined similarly for attributes in $Y$; (4) $R_i.X \cong N$ is the conjunction of not exists (select * from $N$

where $L.cid = N.cid$ and $N.pos = \text{'LHS'}$ and

$L.att = \text{'A'}$ and $R_i.A_j = N.val$)

for each $j \in [1, m_1]$; (5) $R_i.Y \cong N$ is defined similarly, but with $N.pos = \text{'RHS'}$; (6) $X_L$ is the set of following attributes

case when $L.A_j$ is null then ($R_i.A_j \not\in A_{L,j}$) for each $j \in [1, m_1]$; (7) Similarly, $Y_R$ is the set of

case when $R.B_k$ is not null then ($R_i.B_k \not\in B_{R,k}$) for each $k \in [1, m_2]$.

Intuitively, detecting violations of CFD$^p$s is a two-step process. First, query $Q_1^p$ detects single-tuple violations, i.e., the tuples $t'$ in $I_i$ that match the LHS of a CFD$^p$ in $\Sigma_{CFD}^p$, but do not match its RHS. Second, query $Q_2^p$ finds multi-tuple violations, i.e., the tuples $t$ in $I_i$ such that (a) there exists another tuple $t'$ in $I_i$, and $t$ and $t'$ match and agree on the LHS of a CFD$^p$ in $\Sigma_{CFD}^p$, but do not agree on the RHS of the CFD$^p$.

Example 1: Using the coding of Fig. 4, two SQL queries for checking CFD$^p$s $\varphi_2$, $\varphi_3$ and $\varphi_4$ of Fig. 2 are given as follows:

$Q_1^p$: select $R_i \ast$ from $R_i, enc_{L}, L, enc_R, R, enc_\pi \ N$

where $L.cid = R.cid$ and

$(L.sale is null or R_i.sale = L.sale or L.sale = \prime_{\prime})$ and

not exists (select * from $N$

where $N.cid = L.cid$ and $N.pos = \text{'LHS'}$ and

$L.att = \text{'value'}$ and $R_i.sale = N.val$)

for each $j \in [1, m_1]$; (3) $R_i.Y \cong R$ is defined similarly for attributes in $Y$; (4) $R_i.X \cong N$ is the conjunction of not exists (select * from $N$

where $N.cid = R.cid$ and $N.pos = \text{'RHS'}$ and

$L.att = \text{'shipping'}$ and $R_i.sale = N.val$)

for each $j \in [1, m_1]$; (5) $R_i.Y \cong N$ is defined similarly, but with $N.pos = \text{'RHS'}$; (6) $X_L$ is the set of following attributes

case when $L.A_j$ is null then ($R_i.A_j \not\in A_{L,j}$) for each $j \in [1, m_1]$; (7) Similarly, $Y_R$ is the set of

case when $R.B_k$ is not null then ($R_i.B_k \not\in B_{R,k}$) for each $k \in [1, m_2]$.

Intuitively, detecting violations of CFD$^p$s is a two-step process. First, query $Q_1^p$ detects single-tuple violations, i.e., the tuples $t'$ in $I_i$ that match the LHS of a CFD$^p$ in $\Sigma_{CFD}^p$, but do not match its RHS. Second, query $Q_2^p$ finds multi-tuple violations, i.e., the tuples $t$ in $I_i$ such that (a) there exists another tuple $t'$ in $I_i$, and $t$ and $t'$ match and agree on the LHS of a CFD$^p$ in $\Sigma_{CFD}^p$, but do not agree on the RHS of the CFD$^p$.
Ma is supported in part by 973 program 2014CB340300 and NSFC 6132207, Fan is supported in part by NSFC 61133002, 973 Program 2012CB316200 and 2014CB340302, ERC-2014-AdG 652976, Guangdong Innovative Research Team Program 2011D005, Shenzhen Peacock Program 1105100030384361, EPSRC EP/J015377/1 and EP/M025268/1, NSF III 1302212, and a Google Faculty Research Award.

Shuai Ma is a professor at the School of Computer Science and Engineering, Beihang University. He obtained his PhD degrees from University of Edinburgh in 2010, and from Peking University in 2004, respectively. He was a post-doctoral research fellow in the database group, University of Edinburgh, a summer intern at Bell labs, Murray Hill, USA, in the summer of 2008, and a visiting researcher of MRSA in 2012. He is a recipient of the Best Paper Award for VLDB 2010 and the Best Challenge Paper Award for WISE 2013. His current research interests include database theory and systems, social data analysis and data intensive computing.

Liang Duan is a PhD student in the School of Computer Science and Engineering, Beihang University, co-supervised by Prof. Jinfeng Huai and Prof. Shuai Ma. He received his MS degree in computer software and theory from Yunnan University in 2014, and BS degree in computer science and technology from Beihang University in 2009. His current research interests include databases and social data analysis.

Wenfei Fan is Professor (Chair) of Web Data Management in the School of Informatics, University of Edinburgh, UK. He is a Fellow of ACM, a Fellow of the Royal Society of Edinburgh, UK, a National Professor of the Thousand-Talent Program and a Yangtze River Scholar, China. He received his PhD degree from the University of Pennsylvania. He is a recipient of the Runner-up for Best Paper Award of ICDE 2014, the Alberto O. Mendelzon Test-of-Time Award of ACM PODS 2010, the Best Paper Award for VLDB 2010, the Roger Needham Award in 2008 (UK), the Best Paper Award for ICDE 2007, the Best Paper of the Year Award for Computer Networks in 2002, and the Career Award in 2001 (USA). His current research interests include database theory and systems.

Chunming Hu is an associate professor at the School of Computer Science and Engineering, Beihang University. He received his PhD degree from Beihang University in 2006. His current research interests include distributed systems, system virtualization, large scale data management and processing systems.

Wenguang Chen is an associate professor at the Department of Computer Science, Peking University. He obtained his PhD degree from Peking University in 2009. He was a visiting scholar at University of Alberta from 2011 to 2012 and at University of Hawaii at Manoa in from 2012 to 2013. His current research interests include data management, data quality and intelligent HCI.