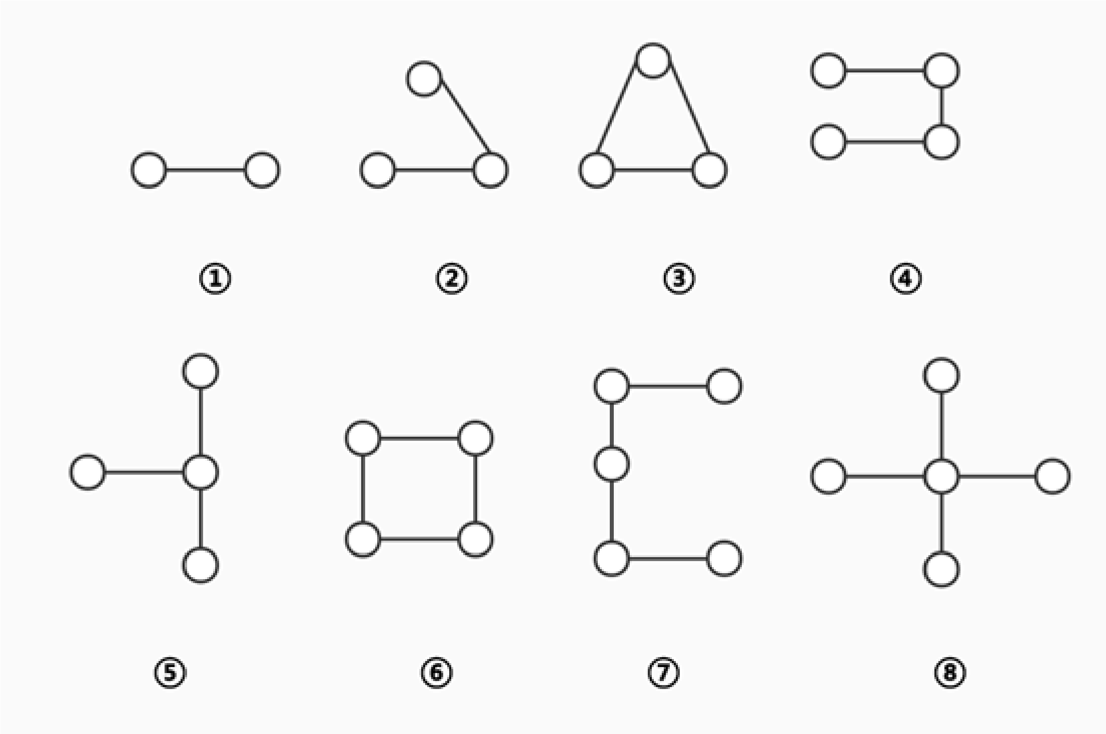
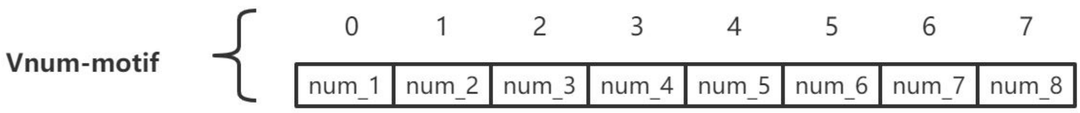
基于Motif Entropy的WL-Kernel方案

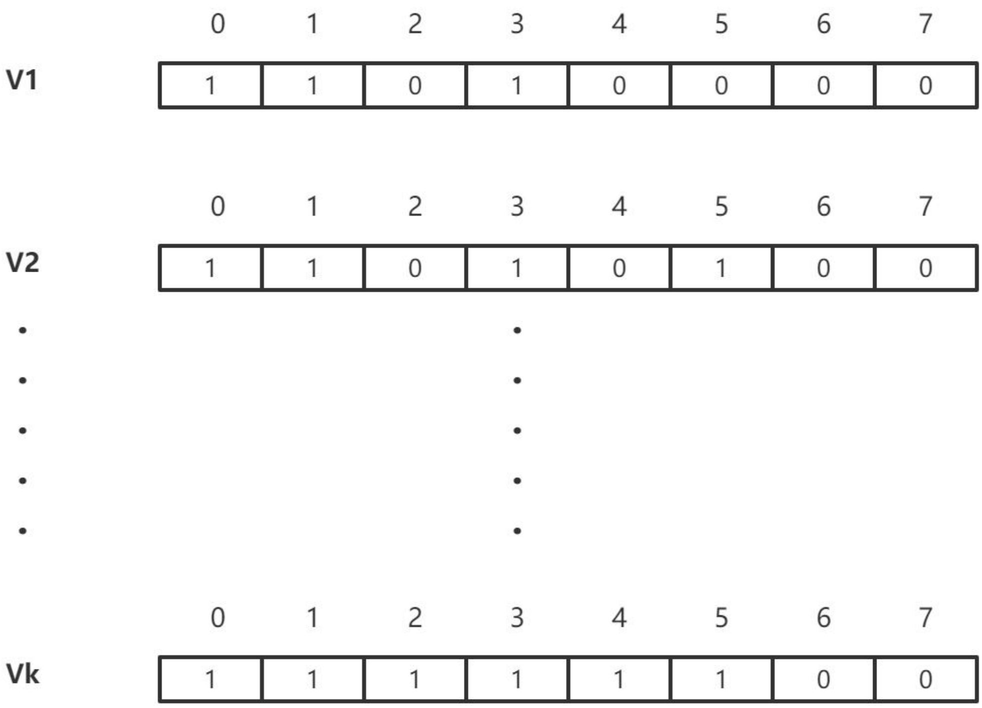


首先给定8种Motif的实例，图$G=(V,E)$，对于节点$n=\{n_1,n_2,...,n_k\}$中的每个节点，分别对k个节点进行度排序，得到排序后从小到大的序列$n'=\{n'_1,n'_2,...,n'_k\}$。

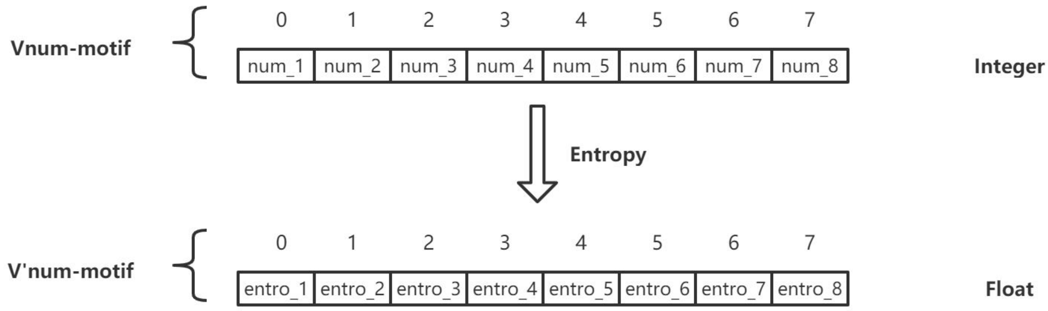
构造一个8维的向量$\mathbf{V_{num-motif}}$，标号0-7，分别对应图中8种Motif，通过对序列n'依次计算Motif出现的次数得出8种Motif的总次数填入向量中，如图所示：



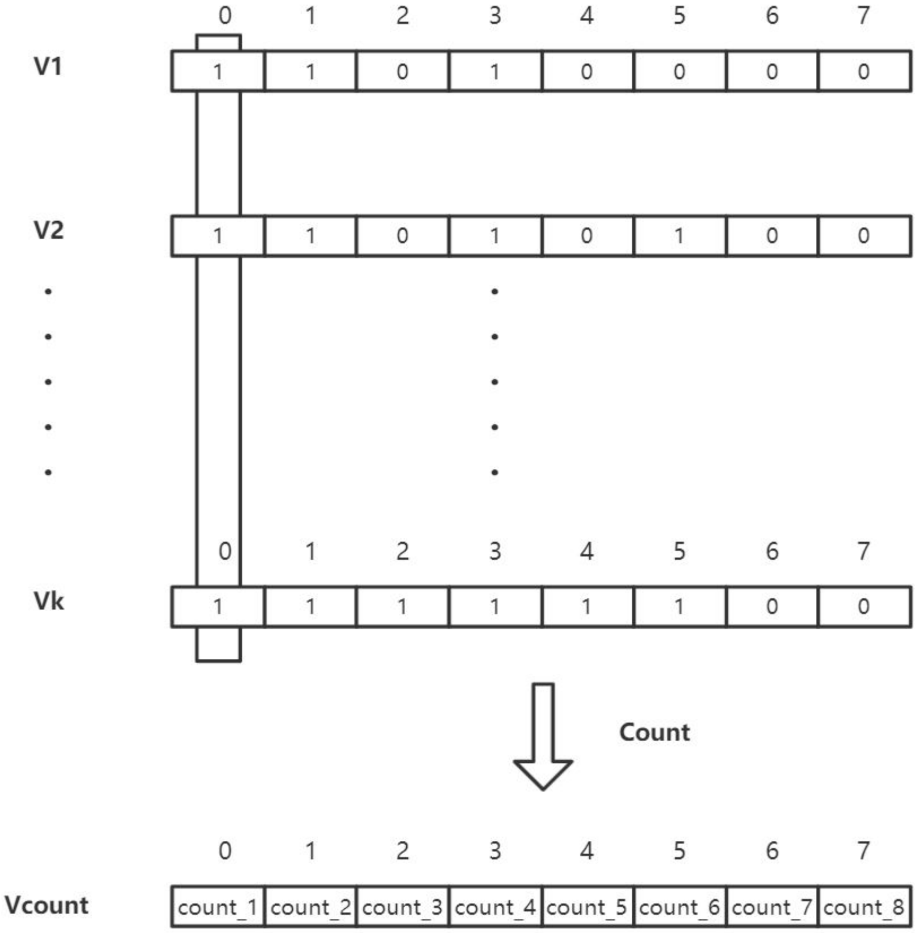
构造k个8维的向量$\mathbf{V_1},\mathbf{V_2},...,\mathbf{V_k}$，分别对应序列$n'$中的节点$n'_1,n'_2,...,n'_k$。同理对每个维度标号0-7，分别对应8中Motif，对于向量$\mathbf{V_1}$，统计节点n'_1是否出现在8种Motif中，出现标记为1，否则为0，得到一个Multi-Hot的8维向量。同理可以得到剩余$k-1$个向量的表达，如图所示：



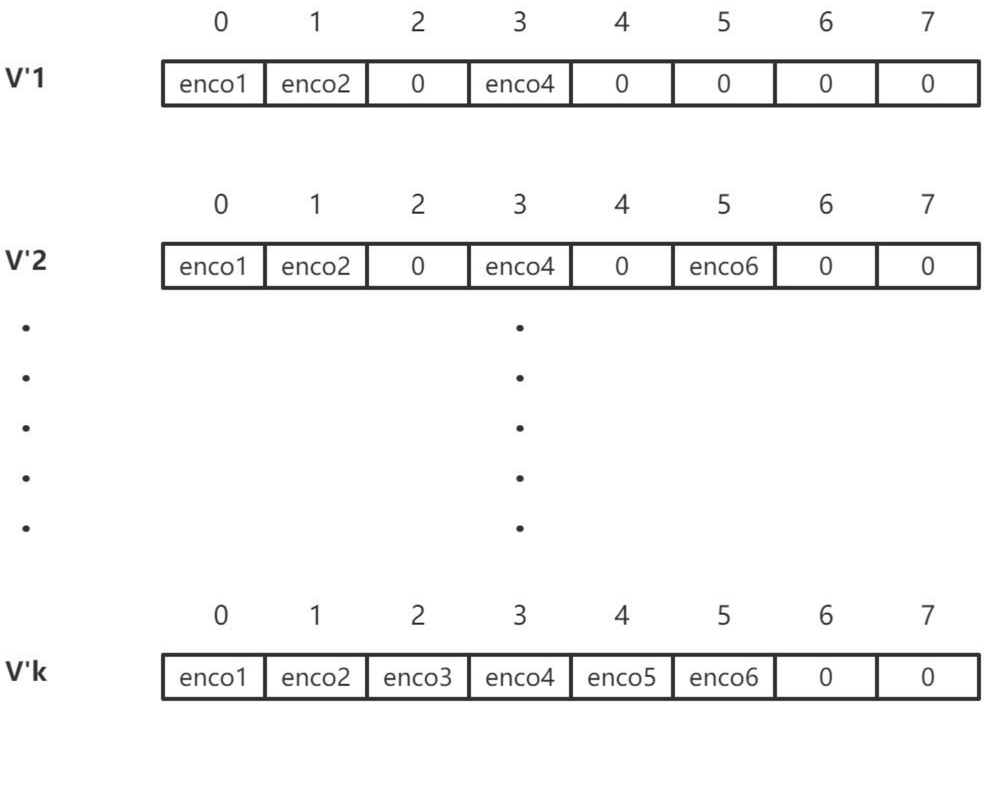
对向量$\mathbf{V}_{\mathbf{num-motif}}$进行entropy运算，得到Motif的entropy表示的熵向量$\mathbf{V}'_{\mathbf{num-motif}}$，如图所示：



对向量$\mathbf{V_1}, \mathbf{V_2}, \ldots, \mathbf{V_k}$的八列分别求和得到向量$\mathbf{V}_{\mathbf{count}}$，$\mathbf{V}_{\mathbf{count}}$的每一列代表Motif的节点出现的总次数，分别记为$Count = \{count_1, count_2, \ldots, count_8\}$，如图所示：



对于熵向量$\mathbf{V}'_{\mathbf{num-motif}}$，分别把对应列0-7与向量$\mathbf{V}_{\mathbf{count}}$的$count_i$的值相除，记为$entropy_{count} = \{enco_1, enco_2, \ldots, enco_k\}$，其中$enco_i = \dfrac{entro_i}{count_i}$，然后依次分配到对应列值为1的向量$\mathbf{V}_{\mathbf{k}}$中得到新的向量$\mathbf{V}'_{\mathbf{k}}$，如图所示：



接下来是对WL-Kernel通过随机游走得到概率分布的方法的改进，为了区分同一标签的节点，我们给其赋予不同的概率，概率分布是通过先前得到的全局的向量$\mathbf{V'_k}$进行局部的概率计算。假设分为m类标签，其中$\mathbf{V'_1}, \mathbf{V'_3}, \mathbf{V'_4}$属于标签$a$类，那么我们分别将$\mathbf{V'_1}, \mathbf{V'_3}, \mathbf{V'_4}$每一列求和分别得到每个向量对应的一个熵值和，记为$entropy^a_{sum} = \{sum^a_1, sum^a_3, sum^a_4\}$，表示$a$类标签下对应向量$\mathbf{V_i}$的熵值和$sum^a_i$。记$a$类标签节点$i$概率为$P^a_i$，则$P^a_1 = \dfrac{sum^a_1}{sum^a_1 + sum^a_3 + sum^a_4}$，可归纳为$P^a_i = \dfrac{sum^a_i}{\sum_j sum^a_j}$。